

VOL. 64, NO. 1

JANUARY 2001

THE MODERN LAW REVIEW

ISSN 0026-9583

PUBLISHED FOR THE MODERN LAW REVIEW LIMITED
BY BLACKWELL PUBLISHERS, OXFORD, UK AND BOSTON, USA

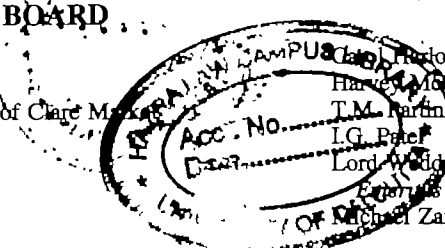
THE MODERN LAW REVIEW

Volume 64

EDITORIAL BOARD

J.S. Anderson
M.R. Chesterman
Lord Dahrendorf of Claremont
Gerald Dworkin
J.M. Evans
Anthony Giddens
Cyril Glasser
J.A.G. Griffith

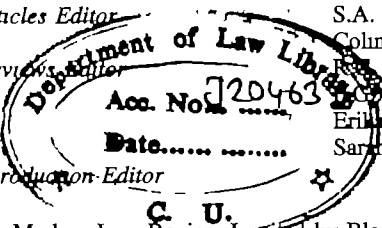
John Frow
Harriet McGregor
T.M. Hartington
I.G. Patel
Lord Wedderburn of Charlton, *Editor*
Editor
OF Michael Zander



EDITORIAL COMMITTEE

Julia Black
Roger Brownsword
Hugh Collins, *Articles Editor*
W.R. Cornish
Neil Duxbury, *Reviews Editor*
Judith Freedman
Nicola Lacey
Martin Loughlin
Kate Maleson, *Production Editor*

W.T. Murphy, *General Editor*
R.W. Rawlings
S.A. Roberts
Colin Scott, *Case Note Editor*
Rex Simpson, *Legislation and Reports Editor*
C. Snyder
Erika Szyszczak
Sarah Worthington, *Secretary*



Published for The Modern Law Review Limited by Blackwell Publishers.

© The Modern Law Review Limited 2001 ISSN 0026-7961

The Modern Law Review Ltd is incorporated in England as a company limited by guarantee under the Companies Act (Reg No 325 282) and is registered as a charity (Reg No 264072)

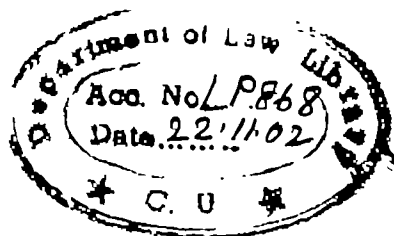
The Modern Law Review is published six times a year, in January, March, May, July, September and November by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main St, Malden, MA 02148, USA.

Information for contributors, subscription information and details about microfilm and advertising appear on the inside back cover.

The Modern Law Review website is at <http://www.lse.ac.uk/depts/law/mlr.htm>

Printed and bound in Great Britain by The Alden Press, Oxford
This journal is printed on acid-free paper.

COPYRIGHT. All rights reserved. Apart from fair dealing for the purposes of research or private study, or criticism or review, as permitted under the UK Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing of the Publisher, or in accordance with the terms of photocopying licences issued by organisations authorised by the Publisher to administer reprographic rights, such as the Copyright Licensing Agency (CLA) and the Copyright Clearance Center (CCC). Institutions with a paid subscription to this journal may make photocopies for teaching purposes free of charge provided such copies are not resold.



LP868

CONTENTS



ARTICLES

The Concept of the Supremacy of the Constitution 1
Jutta Limbach

Crime in a Convict Republic 11
John Braithwaite

The Wasted Costs Jurisdiction 51
Hugh Evans

LEGISLATION

Democracy by Default: The Representation of the People Act 2000 63
Heather Lardy

REPORTS

The Royal Commission on Reform of the House of Lords: A House for the Future? 82
Meg Russell and Richard Cornes

CASES

Increasing the Price of Pain: Damages, The Law Commission and *Heil v Rankin* 100
Richard Lewis

Public Law Representations and Substantive Legitimate Expectations 112
Melanie Roberts

REVIEW ARTICLE

The Developing International Law of Democracy 123
Richard Burchill

53

REVIEWS

Daintith and Page: The Executive in the Constitution: Structure, Autonomy, and Internal Control 135

Walters and Davis-White: Directors' Disqualification: Law and Practice 139

Collier and Lowe: The Settlement of International Disputes: Institutions and Procedures 140

Backhaus: The Elgar Companion to Law and Economics 142

Bouckaert and de Geest: Encyclopedia of Law and Economics. Vols I-V 142

Ewing and Gearty: The Struggle for Civil Liberties: Political Freedom and the Rule of Law in Britain, 1914-1945 143

Dine: The Governance of Corporate Groups 145

Collins: Regulating Contracts 147

Karpik: French Lawyers A Study in Collective Action, 1274-1994. 148

Aust: Modern Treaty Law and Practice 149

Douzinas: The End of Human Rights: Critical Legal Thought at the Turn of the Century	151
Cotterrell: Emile Durkheim: Law in a Moral Domain	152
Revesz, Sands and Stewarts (eds): Environmental Law, the Economy, and Sustainable Development	154

THE MODERN LAW REVIEW

Volume 64

January 2001

No 1

The Concept of the Supremacy of the Constitution

*Jutta Limbach**

The article compares and assesses the idea of the supremacy of the constitution found in Germany with the competing British tradition of parliamentary sovereignty. It concludes by examining the need for a supreme constitutional law in the European Union.

The concept of the supremacy of the constitution confers the highest authority in a legal system on the constitution. Stating this principle does not mean just giving a rank order of legal norms. The point is not solely a conflict of norms of differing dignity. The principle of the supremacy of the constitution also concerns the institutional structure of the organs of State. The scope of the principle becomes clear if we reformulate it: the supremacy of the constitution means the *lower ranking* of statute; and that at the same time implies the *lower ranking* of the legislator.¹

For the attentive British reader, the immediate association will be the contrary principle of parliamentary supremacy or sovereignty, which is a salient feature of English constitutional law. This principle of parliamentary sovereignty means – according to Dicey's definition – that Parliament 'has, under the English constitution, the right to make or unmake any law whatever; and further, that no person or body is recognised by the law of England as having a right to override or set aside the legislation of Parliament'.² Dicey summed up this doctrine in a 'grotesque expression which has become almost proverbial': 'It is a fundamental principle of English lawyers, that Parliament can do everything but make a woman a man, and a man a woman'.³

As we all know, time's gnawing tooth has chipped away even at this principle; or more exactly, European integration has. Yet in its archetypal exaggeration the principle is well suited for legal comparison with the doctrine that marks the German constitution, the supremacy of the constitution. The principle's practical consequence may most easily be seen from a concrete example.

In 1957 the Bundestag enacted a law reordering family law in accordance with the constitutional requirement of sex equality. The Equal Rights Act removed the

* President of the Federal Constitutional Court, Federal Republic of Germany. This is the text of the twenty-eighth Chordey Lecture, delivered at the London School of Economics and Political Science on 31 May 2000.

1 As pertinently put by Rainer Wahl, 'Der Vorrang der Verfassung' *Der Staat* 4/81, 485.

2 A.V. Dicey, *Introduction to the Study of the Law of the Constitution* (8th ed, 1915, Indianapolis: Liberty Fund, 1982) 3ff. Dicey makes clear that, as a legal term, 'Parliament' means 'the King, the House of Lords, and the House of Commons'.

3 *Ibid.*, Dicey quoting De Lolme.

husband's right of decision in matrimonial matters. The Act also followed the case law that parental care and custody go to both father and mother. But in the case where the parents were unable to agree, the father was to have the last word. Only this parental casting vote was held suitable to safeguard family peace and marriage in its Christian, Occidental form. Parliament felt it could deduce from the natural difference of the sexes which parent was due the casting vote. Opinion survey findings from the nineteen fifties show that at that time there was not yet any clear majority view on the question among the population. The paternal right to final decision still had numerous supporters – especially among men.⁴

The Act had barely entered into force when four married mothers filed a constitutional complaint with the Federal Constitutional Court. They asked for repeal of the law giving fathers the last word on child-rearing, because it infringed Article 3 (3) Basic Law. This article of the German constitution reads: 'Men and women shall have equal rights'. The Federal Constitutional Court declared the paternal right of final decision null and void, a year after the Act came into force. It did so irrespective of the fact that the model of an equal footing for man and woman had not yet fully taken over in the legal reality. The Court was unable to see how far objective biological or functional differences or the special nature of woman could justify the paternal prerogative.⁵

The Federal Constitutional Court thus repealed a provision enacted by a majority of the legislators elected by the people. The judges disregarded majority rule. Yet the judges are neither elected by the people, nor owe them any responsibility. For they cannot be removed through new elections, and thus cannot be called to account.⁶ Is this form of judicial review not deeply undemocratic?

We usually deny this by referring to the supremacy of the constitution. This principle in its mature form is a product of American constitutional legal thinking. No one has put it in stronger words than one of the fathers of the American federal Constitution, namely Alexander Hamilton. He wrote in the *Federalist Papers*:

There is no position which depends on clearer principles, than that every act of a delegated authority, contrary to the tenor of the commission under which it is exercised, is void. No legislative act therefore contrary to the constitution can be valid. To deny this would be to affirm that the deputy is greater than his principal; that the servant is above his master; that the representatives of the people are superior to the people themselves; that men acting by virtue of powers may do not only what their powers do not authorise, but what they forbid.⁷

The principle of the legal supremacy of the constitution was then explicitly formulated for the first time in Art 6 of the US Constitution, as follows:

[The] constitution ... shall be the supreme law of the land; and the Judges in every State shall be bound thereby, anything in the Constitution or Laws of any State to the Contrary notwithstanding.

This notion of the constitution as paramount law made the nullity of unconstitutional acts conceivable. This theoretical perception was provoked by

4 R. König, 'Familie und Autorität: Der deutsche Vater im Jahre 1955' in R. König, *Materialien zur Soziologie der Familie* (2nd ed, Köln: Kiepenheuer & Witsch, 1974) 214, 224 ff, and Fröhner, Stackelberger and Eser, *Familie und Ehe* (Bielefeld: Stackelberg, 1956)

5 BVerfGE 10, 59 <74 f.>.

6 Ronald Dworkin, "Gleichheit, Demokratie und die Verfassung. Wir, das Volk und die Richter", in *UK Preuß, Zum Begriff der Verfassung* (Frankfurt am Main: Fischer Taschenbuch Verlag, 1994) 171.

7 Alexander Hamilton, *The Federalist* No. 78, in Gary Wills (ed), *The Federalist Papers by Alexander Hamilton, James Madison and John Jay* (New York: Bantam Books, 1982) 395.

the unfortunate experience of a legislature that saw itself as sovereign.⁸ The experience of the North American colonies had been that a Parliament can also do injustice through the repeated attempts by the British Parliament, without any involvement of the colonists' assemblies, to tax articles of everyday necessity in the colonies. The assemblies declared the Acts unconstitutional and insisted on the right of all British citizens to be taxed only on the basis of a law in which they had been involved at least through elected representatives: 'No taxation without representation'.

These revolutionary events, to which we can only allude here, led to a fundamental rethinking of the problem of the limits to State power. This culminated in the image of a State order that is both democratic and guarantees the individual's fundamental rights.⁹ This is the principled answer to the accusation that judicial review is incompatible with the democratic principle. Majority rule does not by itself constitute the essence of Western-style democracy. Democracy means not just that State power derives from the people and politics is determined by their elected representatives. Another part of democracy comprises particular fundamental values, to which all organs of State are committed. It is not only the two dictatorships on German soil that have taught us that democracy cannot be upheld without the validity of human rights. 'Democracy is thus a delicate balance between majority rule and particular fundamental values, such as human rights'.¹⁰

The judges' criterion is the Basic Law and not public opinion or what the majority of the population think – however desirable society's acceptance of the judgments may be. Thus the judges have to '*defy the general will* as soon as constitutional guarantees are at stake'.¹¹ Should, however, the *protection of a minority* be concerned, then concern at public protest or public conflict cannot oblige the Court to remain reticent. This advantage of judicial activity may be one of the reasons why politics is sometimes very ready to leave unpopular decisions to the Federal Constitutional Court.

Instead of pursuing the concrete development of the primacy of the constitution through American constitutional history, let us use the German constitution as an example. For the framers of the Basic Law in 1948–9 transformed American experience and achievements into written constitutional law. This also applies to judicial review, which was born with the US Supreme Court's judgment in *Marbury v Madison*.¹²

There are three traits that primarily characterise the principle of supremacy of the constitution: 1. The possibility of distinguishing between constitutional and other laws; 2. the legislator's being bound by the constitutional law, which presupposes special procedures for amending constitutional law; and 3. an institution with the authority in the event of conflict to check the constitutionality of governmental legal acts.¹³

The Basic Law opens with a catalogue of human rights, crowned by the principle of human dignity. The failure of the Weimar Republic and the experience of the totalitarian Nazi regime did not only induce the framers of our constitution to cast

8 As well put by Wahl, n 1 above, 489 f

9 *Ibid* 490.

10 Aharon Barak, *Judicial Discretion* (New Haven: Yale University Press, 1989) 125

11 Ronald Dworkin, n 6 above, 172. See also Eric Barendt, *An Introduction to Constitutional Law* (New York: Oxford University Press, 1998) 24.

12 1 Cranch 137 (1803).

13 cf n 1 above, 490f, who requires as a further prerequisite that the constitution be understood as a self-binding of the sovereign people in favour of the individual's human rights. cf also n 2 above, 40, though here speaking of features of a 'non-sovereign law-making body'.

human rights and freedoms as enforceable subjective rights. Additionally, they allotted the fundamental rights a special rank: the Basic Law explicitly orders that the basic rights shall bind the legislature, the executive and the judiciary as directly applicable law (Article 3 (3) Basic Law). As Dieter Grimm rightly emphasised: 'It was the firm consensus of all political forces active in the constitutional assembly to prevent another failure of representative democracy in Germany and to establish effective safeguards against dictatorship and disregard of human rights. The constitution should therefore be the paramount law of the land and claim priority over any government act'.¹⁴

It is not only the basic rights that limit the right of the majority to decide as it pleases. There are special provisions in our Constitution for its own amendment. Moreover, basic constitutional principles which structure the State – such as democracy, separation of powers, the rule of law and respect for human dignity – may not be altered at all. The aim was to prevent the enemies of democracy from overturning it using its own instruments – like majority rule.

To have laid down the principle of supremacy of the constitution with distinctness is much, as Dicey says, 'but the great problem was how to ensure that the principle should be obeyed'.¹⁵ As we all know, a Bill of Rights, embodied in the constitution, is not by itself sufficient to establish a stable democracy committed to fundamental values and the rule of law. Democracy must be continually cherished and defended. The challenge must be met not only by those engaged in politics as a profession. Accordingly, in the German system all constitutional organs have to respect and enforce the human rights. This duty has to be observed by all administrative bodies, and by the judiciary.

But the principle of supremacy of the constitution attains its practical point only with the introduction of constitutional jurisdiction.¹⁶ The authors of the Basic Law established a special institution designed to enforce the constitution against any other government authority. It was also to be a safeguard against dictatorship and disregard of human rights. This institution is the Federal Constitutional Court, endowed with ample powers, among them judicial review of legislation as well as of executive acts and judicial decisions.

Only this Court may find that a law is incompatible with the Basic Law. Should another court consider a law to be unconstitutional and therefore wish not to apply it, it must first obtain the decision of the Federal Constitutional Court.¹⁷ This Court is not a general court of review which examines the decisions of ordinary courts for any error of fact or law. Its exclusive responsibility is to decide questions of constitutional law and to interpret and apply the Basic Law with final binding force.¹⁸ The Court enjoys the last word on the meaning of the Basic Law, and its word is law.¹⁹

14 Dieter Grimm, 'Human Rights and Judicial Review in Germany' in David M. Beatty, (ed), *Human Rights and Judicial Review, A Comparative Perspective* (Dordrecht: Nijhoff, 1994) 267–295, 270.

15 n 2 above, 88.

16 n 1 above, 485.

17 Additionally, the Federal Government, a State Government or one third of the members of the Bundestag may have the constitutionality of a legal norm reviewed. We speak with regard to this kind of proceedings of abstract norm control.

18 See Art 31(1) of the Federal Constitutional Court Act: 'The decisions of the Federal Constitutional Court shall be binding upon federal and Land constitutional organs as well as all courts and authorities.'

19 See Art 31(2) 'If a law is declared to be compatible or incompatible with the Basic Law or other federal law or to be null and void, the decision shall be published in the Federal Law Gazette by the Federal Minister of Justice, because these decisions shall have the force of law.'

In recent history, the concept of constitutional jurisdiction has become, in Europe and globally, a consistent feature of democratic governance. At present, this no longer applies only to Europe. Constitutional courts also operate, for example, in South Korea, in Mongolia, in the Republic of South Africa, and in most of the countries of Central and South America. The widespread appearance of constitutional jurisdiction in Western European countries following World War II was a clear reaction to authoritarian and totalitarian forms of government. This institution was perceived as a necessary guarantee of a democratic legal system. The mass emergence of constitutional courts was also a characteristic feature of the system transformations in Central and Eastern Europe at the turn of the 1980s, in the countries of so-called 'real socialism'. By this means political reforms were often started, and the constitutional courts became their guarantor. In this development, the German system served as a model, in the same way as the authors of the Basic Law looked to the US Supreme Court as a model for establishing a special court of constitutional review.

The power of judicial review is a chief legal instrument in the system of checks and balances. Nevertheless, it is not a universal or necessary element of a democratic constitution. We know some democracies that have neither constitutional jurisdiction nor judicial review in the strict sense, although – or just because – their constitution is firmly based on the separation of powers. Consider the Scandinavian countries, or the Netherlands. Britain in particular had no judicial review until its membership of the European Union. In 1915 Dicey was still able to say imperturbably: 'No British Court can give judgment, or ever does give judgment, that an Act of Parliament need not be obeyed because it is unconstitutional',²⁰ because the doctrine of parliamentary supremacy is 'the very keystone of the law of the [British] constitution'.²¹

The dominance of this principle in Britain likewise has historical reasons. The doctrine of fundamental rights has – as we all know – important British sources. Consider Magna Carta of 1215, the Petition of Rights of 1628, the Habeas Corpus Act of 1679 and the Bill of Rights of 1689. In early 17th century Britain there were even delicate beginnings of the idea of the primacy of fundamental rights and principles.²² Yet this idea 'became eclipsed at the end of the seventeenth century by the concept of absolute Parliamentary sovereignty,' as Anthony Lester and David Pannick put it.²³ While judges had at the time fought both for their independence and for judicial review, they won only the fight for judicial independence. That was more or less the price they had to pay for their alliance with Parliament against the Crown. In the words of Lester and Pannick, 'The "glorious bloodless" revolution was won by Parliament; and although the Bill of Rights of 1688–89 and the Act of Settlement of 1700 recognised some important personal rights and liberties, the terms of the constitutional settlement were mainly concerned with the rights and liberties of Parliament. The alliance of Parliament

20 n 2 above, 47.

21 *ibid* 25.

22 One of England's most famous judges, Sir Edward Coke, stated in a judgment in 1610 (in the case of Dr Thomas Vonham) that if a law run counter to common right and reason, then it is contradictory and therefore unenforceable. He pronounced it null, in the words: 'the common law will control it and adjudge such act to be void.' Reported in Gerald Stourzh, *Vom Widerstandsrecht zur Verfassungsgerichtsbarkeit: Zum Problem der Verfassungswidrigkeit im 18. Jahrhundert* (Graz: Styria, 1974) 14.

23 Anthony Lester and David Pannick (eds), *Human Rights – Law and Practice* (London: Butterworths, 1999) 1.

and the common lawyers ensured that the supremacy of the law would mean the supremacy of Parliament'.²⁴

It would, though, be wrong to believe that fundamental values played no part in the case law of states without constitutional jurisdiction. Each democratic judiciary will develop tools to perform judicial control of legislative acts, for example by interpretation in the light of a fundamental principle. Thus, the British courts apply a kind of implicit constitution-conformable interpretation. This means that if the wording of the statute so allows, the court will interpret parliamentary legislation in such a way as to avoid a violation of fundamental principles.²⁵ Since the passage of the European Communities Act 1972 the British courts play 'a key role by seeking, wherever possible, to interpret national law consistently with Community law, thus ensuring that the latter takes effect throughout the Union'. So said the Lord Chancellor at the Millennium Lectures a few weeks ago.²⁶

Moreover, the House of Lords decided that parliamentary legislation should not be applied when it is incompatible with a provision of Community law. The grounds of the House of Lords ruling state that Parliament 'had voluntarily accepted a limit on its legislative powers through the passage of the European Communities Act 1972',²⁷ in which Parliament directed the courts to give priority to directly effective Community law. While this is not the end of the principle of parliamentary supremacy, it is nonetheless a not inconsiderable limitation of it.

This experience associated with European integration has not managed to shake British loyalty to the principle of parliamentary supremacy. On the contrary, it strongly influenced the debate as to how – and with what judicial powers – the European Convention on Human Rights ought to be incorporated into national law.²⁸ The key provision of the Human Rights Act imposes a duty upon all public authorities to act compatibly with Convention rights. But just as before, British courts are not entitled to set aside parliamentary legislation on the ground that it violates human rights.

The Human Rights Act 1998 does not alter this position. It lays down the following procedure. First, courts have to interpret national legislation, wherever possible, in a way which is compatible with Convention rights. Where this is not possible, the higher courts have to make a formal declaration that national legislation is incompatible with a Convention right. Such a declaration of incompatibility does not affect the validity, continuing operation or enforcement of the provision in question, nor is it binding on the parties to the proceedings in which it is made. The declaration merely sets in motion a fast-track procedure to amend incompatible legislation.²⁹ Thus, the power to amend parliamentary legislation is strictly reserved to Parliament itself and – under certain provisions³⁰ – to the government. The same rules apply if the European Court of Human Rights states the incompatibility of British national law with the Convention.

24 The sentence continues: 'more realistically, the supremacy of the central government in Parliament, Lord Hailsham of St Marylebone's "elective dictatorship"'. *Ibid.*

25 Rob Bakker, "Verfassungskonforme Auslegung" in Rob Bakker et al (eds), *Judicial Control – Comparative Essays on Judicial Review* (Antwerpen: Maklu, 1995) 20.

26 Lord Irvine of Lairg, 'The Influence of Europe on Public Law in the United Kingdom' in Basil Markesinis (ed), *The Clifford Chance Millennium Lectures: The Coming Together of the Common Law and the Civil Law*, (Oxford: Hart, 2000) 11, 12.

27 n 11 above 99.

28 See n 23 above, 11–12, and n 26 above, 14–16.

29 Human Rights Act 1998, s 3(1), ss 4(2) and (6).

30 Human Rights Act 1998, s 10.

This scheme, which has its roots in the New Zealand Bill of Rights Act 1990, is described by Anthony Lester as 'an ingenious and successful reconciliation of the principles of parliamentary sovereignty and the need for effective domestic remedies'.³¹ Whether this mode of proceeding is in fact the philosopher's stone is something we shall learn from coming years of British court practice. Comparative lawyers and legal sociologists will find a research field here that is as interesting as it is important. For not just in Britain, but in all countries with a Bill of Rights, it is a thorny question how to draw a line between the law-making power of the legislature and judicial review.

Allow me to bring this out by looking at the key problem the concept of primacy of the constitution brings, as illustrated by German constitutional law. It should first be stated that the Federal Constitutional Court performs a predominantly checking function, delimiting and restraining power. Its job is to tie policy to law, and subordinate it to law. For the Basic Law has 'resolved the age-old tension between power and law in favour of the law'.³² The Court's task is the limited one of interpreting the Constitution in Court proceedings; considerations of expediency found in politics are not its concern.³³ The Federal Constitutional Court is 'designed as an organ of law, not of politics'; even if its decision may, inevitably, have political repercussions.³⁴ Nonetheless it is a crucial question whether constitutional adjudication can at all be separated with logical distinctness from lawmaking.

This difficulty has to do with the specific nature of the text. The Articles of the Basic Law are marked by a low degree of definiteness. The Basic Law, considered from a structural viewpoint, is a framework order (*Rahmenordnung*). With very few exceptions, it formulates no directly applicable provisions. Its articles are norms with great openness, and margins of interpretation that are hard to delimit.³⁵ The Basic Law essentially contains – apart from the law on the organisation of the State – principles that must first be spelled out before they can be applied.³⁶

Let us take as an example an important article devoted to the family. It opens with the sentence: 'Marriage and the family shall enjoy the special protection of the State.' (Article 6(1) Basic Law). What is meant by the term 'family'? Only a family founded on marriage? That was the opinion of the framers of the Basic Law. Or should we understand 'family' in the light of present-day circumstances, and interpret the term in line with the sociology of the family? That was and is a recurrent dispute not just among German constitutional scholars. The Federal Constitutional Court too has repeatedly dealt with, for instance, the question

31 n 26 above, 15.

32 Helmut Simon, "Verfassungsgarantbarkeit" in E. Benda, W. Maihofer and H. Vogel (eds), *Handbuch des Verfassungsrechts* (Berlin/New York: de Gruyter, 2nd ed, 1994) 1137 ff, 1661.

33 So the Federal Constitutional Court refrains from considering whether the legislature has chosen the wisest, most just, and most expedient solution. Cf BVerfGE 36, 174 <189>, 38, 312 <322>.

34 As rightly put by Thomas Clemens, 'Das Bundesverfassungsgericht im Rechts- und Verfassungsstaat: Sein Verhältnis zur Politik und zum einfachen Recht; Entwicklungslinien seiner Rechtsprechung' in Michael Piazolo (ed), *Das Bundesverfassungsgericht – ein Gericht im Schnittpunkt von Recht und Politik* (Mainz – München: Hase & Kohler, 1995) 13ff, 16f, and Dieter C. Umbach, 'The German Democracy and the Federal Constitutional Court as Promoter and Guardian of the Rule of Law' in *Democracy and the Rule of Law in Germany* (Jordan: Konrad Adenauer Foundation, 1992) 25f.

35 K. Hesse, *Grundzüge des Verfassungsrechts der Bundesrepublik Deutschland* (Heidelberg: Müller, 1995) 20, and Wolfgang Zeidler, *Verfassungsgerichtsbarkeit, Gesetzgebung und politische Führung, Ein Cappenberger Gespräch* (Köln: Grote, 1980) 46.

36 On the foregoing cf Ernst-Wolfgang Böckenförde, "Die Methoden der Verfassungsinterpretation" in Ernst-Wolfgang Böckenförde, *Staat, Verfassung, Demokratie* (2nd edn, Frankfurt a M: Suhrkamp, 1992) 53 ff, 58.

whether there could be a family in the legal sense where father and mother lived together unmarried with their child. At first the Federal Constitutional Court took the view that the child born out of wedlock had a constitutionally protected relationship with each of its father and mother. It was not, however, prepared to treat extramarital cohabitation, with a child, as a family within the meaning of the constitution.³⁷ This distinction, felt to be odd not to say grotesque, could hardly be made comprehensible to the legal layman.

Gradually the insight dawned on the Court that the constitutionally regulated protection of the family could not be made dependent on whether the parents had found their way to the registry office. For what deserves protection is the fact that people live together and collaborate in order to raise and bring up children. The Court gradually adjusted its legal concept of the family to social reality. The legislator too then renewed family law accordingly. Among other things, parents living together unmarried today have the possibility of joint custody.

At present one vigorously argued question is how the constitutional concept of marriage is to be understood. Is it essential to marriage within the meaning of the constitution that it be between partners of opposite sex, or is this state-protected institution also open to homosexual couples? The Federal Constitutional Court has to date confined constitutional protection to a life partnership of man and woman, so that same-sex couples in Germany are still barred from the registry office.³⁸ The institution of marriage could accordingly not even be opened up to same-sex couples by an ordinary Act.³⁹ It would take a constitutional amendment by the requisite qualified majority.⁴⁰ But how would the Federal Constitutional Court decide, if one day the social conception of marriage were to have fundamentally changed? Here the fact that more and more and more children are being brought up in such partnerships might shift social thinking and opinions.

The openness and breadth of the Basic Law sketched out here ought not to be pointed to as defects. On the contrary: a constitution can generally be regarded as successful if it is couched tersely and vaguely. For a constitution that were not open and therefore to some extent capable of ever-new interpretation would inevitably soon come into hopeless contradiction with its object.⁴¹ That is why a constitution has to be understood as a living instrument that has to be interpreted in the light of current circumstances.⁴²

We must accordingly concede that judicial decision making is not only law-finding, but always also law-making. The judge creates law in the process of finding a decision. Adjudication thus always has a political dimension too. This is certainly true of constitutional jurisdiction. But though interpreting the constitution cannot be reduced to textual exegesis, the question where constitutional jurisdiction's area of action ends and that of politics begins demands an answer, however imperfect. For an institution which, like the Federal Constitutional Court, reviews the functioning of the constitutionally set limits must in turn be mindful of the limits to its own decisional power.

To be honest there is no binding constitutional theory or catalogue of useful criteria that could serve as a signpost in the ridge-walking between law and

37 BVerfGE 56, 363 <386>.

38 BVerfG (3rd chamber of the First Senate) in NJW 1993, 3058, and the case law cited there.

39 See Walther Panly, 'Sperrwirkungen des verfassungsrechtlichen Ehebegriffs' in NJW 1997, 1955.

40 Meaning two thirds of the Members of the Bundestag and two thirds of the votes of the Bundesrat: Art 79(2) Basic Law.

41 Willi Geiger, *Verfassungsentwicklung durch das Bundesverfassungsgericht* (Karlsruhe: private manuscript 1965) 4.

42 As the European Court of Human Rights said in relation to the ECHR, in NJW 1999, 3109.

politics.⁴³ Attention is therefore increasingly being paid to theorems that circumscribe a line of thought that takes as its object the relation between constitutional jurisdiction and the other organs of State laid down in the Basic Law.⁴⁴ What could be more appropriate than to look for the information, or better the stimulus to thought, in the constitutional principles whose infringement is continually alleged in cases of boundary disputes? In the forefront here are the principle of the separation of powers and the democratic principle.

The maxims derived from those principles are necessarily very abstract. And they denote only stages in a thought process which has to be paced off in each and every case of constitutional decision-making.⁴⁵ These topics of reflection are very similar to the arguments used in Britain to justify the principle of parliamentary supremacy.

One of the good reasons for the supremacy of Parliament is for Lord Irvine the 'pragmatic imperative'. The courts, he argues, have considerably less expertise than the parliament, particularly on substantive matters of policy. It is therefore desirable 'that the authority [i.e. Parliament] itself should make such decisions because it is better equipped to do so'.⁴⁶ We speak of the functional aspect of the principle of separation of powers. This – according to the Federal Constitutional Court – is aimed primarily at having 'government decisions taken as correctly as possible, that is, by the organs best meeting the requirements for them, because of their organisation, composition, function and procedures'.⁴⁷ British lawyers call this a 'pragmatic imperative'.

Another good reason for parliamentary supremacy is the so-called 'democratic imperative'. This says that the electoral system operates as an important safeguard against misuse of public power by requiring many public authorities to submit themselves to the verdict of the electorate at periodic intervals. In the German system the democratic principle calls for self-restraint by the Federal Constitutional Court.⁴⁸ It is argued that the openness of the political process essential to democracy takes account of the fact that the Basic Law leaves political decision-making up to as broad as possible a political debate, in which the common welfare is defined.⁴⁹ This applies above all in questions where the Basic Law is not very eloquent. For in all areas not adequately provided for in constitutional law, 'the democratic principle requires that only those may decide who can be called to account again through the vote'.⁵⁰

Response by the citizens in regular elections does not apply to the members of the Federal Constitutional Court. The judges are outside the day-to-day political struggle. Their personal and material independence, guaranteed in the Basic Law, is intended to make them immune specifically to political requirements, and

43 Jutta Limbach, 'The Law-Making Power of the Legislature and the Judicial Review' in Basil Markesinis (ed), *Law Making, Law Finding, and Law Shaping* (Oxford: Oxford University Press, 1997) 161.

44 cf eg Brun-Otto Bryde, *Verfassungsentwicklung* (Baden-Baden: Nomos, 1982) 325ff; Grimm, 'Verfassungsgerichtsbarkeit – Funktion und Funktionsgrenzen im demokratischen Staat' in Wolfgang Hoffmann-Riem (ed), *Sozialwissenschaften im Studium des Rechts II, Verfassungs- und Verwaltungsrecht* (Munich: Beck, 1977) 83ff, 98ff; Konrad Heise, 'Funktionelle Grenzen der Verfassungsgerichtsbarkeit' in *Recht als Prozeß und Gefüge, Festschrift für Hans Huber* (Bern: Staempf, 1981) 261ff, and n 32 above, 1665ff.

45 See n 43 above, 169–175.

46 See n 26 above, 22.

47 BVerfGE 68, 1 <86>.

48 n 43 above, 171–174.

49 Bryde, n 44 above, 343.

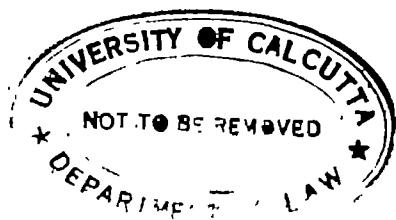
50 Grimm, n 44 above, 100.

guarantee that the law alone counts. To decide on the optimum realisation of the common welfare is by contrast a matter for politics.

The British and the German constitutions are equally committed to the principles of democracy and the separation of powers. All the same, we have chosen different patterns of solutions to the same problems. Neither of these two principles – parliamentary supremacy and the primacy of the constitution – can claim absolute rightness for itself: The decisive question is what principle best suits the relevant legal system, political culture and historical experience. We Germans cannot say with the same pride as one old English lawyer did that we owe the advantage of a long and uninterruptedly increasing prosperity of our country to the spirit of our laws.⁵¹ We have to grapple with a disastrous past, which has told us that a democracy cannot be maintained without the positive validity of fundamental rights.

Yet, despite our differing approaches to solutions, it is fruitful to consider from a comparative law viewpoint how these differing approaches stand up. For any system is susceptible of improvement in the light of new experience and new insights. What is more, the problem of norms that compete with each other is not just a national but also a supranational – i.e. European – one. And the point here is not just conflicts of norms but the rights and duties of European and national authorities.

With the advance of European integration, constitutional questions in the supranational framework are becoming ever more pressing. The Charter of Fundamental Rights should be a first step along this road. European citizens do not want to have to extricate their fundamental rights laboriously out of hundreds of judgments and the chaotic tangle of the treaties. I do not share the concern that the Charter may prove to be a Trojan horse and undermine the sovereignty of Member States. The automatic association of constitution with statehood has become anachronistic and questionable. The historically-grown European nation States will remain the indispensable basis of international and supranational organisations. But just because that is so, there is need of a constitution to confirm and to delimit the power of Brussels. A constitution could at long last provide clarity, and not just about the rights of Union citizens. It could set bounds on Europe, and once and for all settle the tension between European and national law. That is why it is important for us to ponder from a comparative-law viewpoint how we are in future to organise the European Union democratically and in accordance with the principle of separation of powers in some sort of basic law. For it is the prime task of a modern constitution to set bounds on political rule, in order to guarantee the freedom of the citizens.



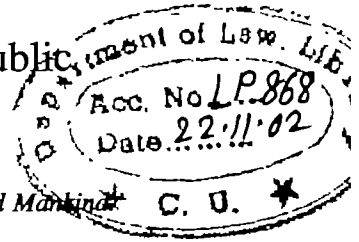
51 As Hallam did in 1818, cited in n 2 above, cxxv

LP 868

Crime in a Convict Republic

John Braithwaite*

'It is much easier to extirpate than to amend Mankind'
Sir William Blackstone

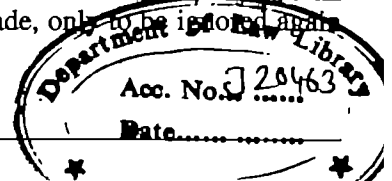


Five stages in the history of regulation are derived from the literature as a starting framework for this essay. These stages are outlined in the first section. This five-stage model is then confronted and revised in light of the neglected case of the Australian penal colony. It is juxtaposed throughout the paper with the history of the regulation of crime in the US. Australian convict society is found to be brutal yet forgiving. We conclude that surprisingly high levels of procedural justice and reintegration in Australian convict society drive down crime rates at a remarkable rate in the nineteenth century. In contrast American slave society is characterised by procedural injustice, exclusion and stigmatisation, which delivers high crime rates. Following Heimer and Straffen's theory, reintegration and procedural fairness are found to arise in conditions where the powerful are dependent on the deviant.¹ Acute labour shortage is the basis of a reintegrative assignment system for Australian convicts to work in the free community. While convicts change Australia in very Australian ways, we find that many of these developments are not uniquely Australian and so a revision of the five-phase model is proposed. The revision also implies that Foucault's distinction between governing the body versus governing the soul (corporal/capital punishment versus the penitentiary) is less central than exclusion versus inclusion (banishment versus restorative justice) to understanding all stages of the history of regulation.

LP 2172

A Perspective on the History of Regulation

The historical vision of criminologists, sociologists and philosophers of punishment alike is impoverished at this point in history. Mainstream thinkers have limited ways of comprehending core problems of theft and violence beyond prescribing the right dose of imprisonment. The most important alternative current, dominated by Michel Foucault's *Discipline and Punish*, is very much a 'history of the present' in the way the rise of the penitentiary is read as the enduring central question rather than a phase!² Rusche and Kirchheimer provided a Marxist history of punishment that was ignored for its first three decades of existence, enjoyed an honoured place in curricula for little more than a decade, only to be ignored again with the passing of Marxism's intellectual heyday.³



* Australian National University.

- 1 Carol A. Heimer and Lisa R. Straffen, 'Interdependence and Reintegrative Social Control: Labelling and Reforming "Inappropriate" Parents in Neonatal Intensive Care Units' (1994) 68 *American Sociological Review* 635-654.
- 2 Michel Foucault, *Discipline and Punish: The Birth of the Prison*, translated by Alan Sheridan (London: Allen Lane, 1977).
- 3 G. Rusche and O. Kirchheimer, *Punishment and Social Structure* (New York: Columbia University Press, 1939).

Part of my intellectual agenda is to de-centre the penitentiary, and indeed punishment, in the history of regulation. In doing so, there is an important place for Rusche and Kirchheimer's materialist analysis,⁴ for the disciplinary analysis of *Discipline and Punish*, and for the governmentality of the late Foucault as well.

Rather standard readings of the sociology of punishment literature are combined to identify a core Western sequence in the history of regulation with five stages:

1. A pre-state stage when restorative justice and banishment are dominant;
2. A weak state stage where corporal and capital punishment dominate;
3. A strong state stage where professional police and penitentiaries dominate;
4. A Keynesian welfare state stage where new therapeutic professions such as social work colonise what becomes probation-prison-parole; and
5. A contemporarily evolving new regulatory state phase of community and corporate policing (with a revived restorative justice).

The first stage in this history of regulatory institutions is a pre-state period lasting to about the 12th century in many European societies. This is the stage currently attracting much attention in the writing of restorative justice scholars.⁵ According to Weitekamp, until the 12th century restorative justice (participatory dialogue oriented to healing rather than hurting) was the dominant form of regulation in pre-state societies, banishment and capital punishment significant back-ups to it.⁶ For many parts of Europe local predominantly kin-based restorative justice dominates the king's punitive justice for four or more centuries beyond the 12th century, for example in Scotland.⁷ Like all divides in the posited sequence, there is much variation in the overlapping of boundaries. Beyond the West, for example in Africa, local restorative justice remains more important than state punishment until and even throughout the 20th century.

The second stage is the period of weak Western states, from 12th to 18th centuries in much of Europe. Weak kings crush indigenous restorative traditions and inflict ever more horrible physical punishments on the bodies of their subjects. It was Michel Foucault who identified this stage and distinguished it from the third stage. Foucault perceptively sees physical punishment as the king inscribing his power on the bodies of subjects, signifying the awe of his rule by highly public forms of humiliation of those who defy it.⁸ These spectacles occur on the scaffold, at the flogging post, the stocks and through branding, for example. Crime is no longer committed against victims:

[C]rime signified an attack upon the sovereign, since the law represented and embodied the sovereign's will. Punishment is thus an act of vengeance, justified by the sovereign's right to make war on his or her enemies and conducted in appropriately warlike terms. In keeping with the military sources of this sovereign power, justice is a manifestation of armed violence, an exercise in terror intended to remind the populace of the unrestrained power behind the law. The body of the condemned here becomes a screen upon which sovereign

4 *ibid.*

5 Daniel Van Ness and Karen Heetderks Strong, *Restoring Justice* (Cincinnati: Anderson, 1997).

6 E. Weitekamp, 'The History of Restorative Justice' in Gordon Bazemore and Lode Walgrave (eds), *Restorative Juvenile Justice* (Monsey, NY: Criminal Justice Press, 1999).

7 Robert E. Mackay, 'The Resuscitation of Assythment? Reparation and the Scottish Criminal' (1992) 3 *Juridical Review* 242-55; Jenny Wormwald, 'Bloodfeud, Kindred and Government in Early Modern Scotland' (1980) 87 *Past and Present* 54-97; Ian D. Whyte, *Scotland Before the Industrial Revolution* (London: Longman, 1995) 217.

8 See also Pieter Spierenburg, *The Spectacle of Suffering: Executions and the Evolution of Repression, from a Preindustrial Metropolis to the European Experience* (Cambridge: Cambridge University Press, 1984).

power is projected, or more precisely a flesh upon which the marks of power can be visibly engraved.⁹

But there is more to the story than Foucault's way of seeing it. Weak kings actively wanted to crush indigenous informal justice because there was political power in centralising prerogatives of mercy into their own hands, more so when punishment was awesome. Hence the demise of restorative justice and the rise of corporal and capital punishment reflect conscious tactics by weak princes to prevent the disintegration of their centralising projects. Foucault sees a shift away from this mode of exercising power between about 1750 and 1820, though Spierenburg's subsequent work shows a more gradual abandonment of spectacles of corporal punishment in Europe between 1600 and the early twentieth century.¹⁰

The third stage is the rise of a strong central state which pursues consistency in the administration of punishment. For most of Europe this occurs throughout the nineteenth century. Oliver MacDonagh's history of the Passenger Acts is the seminal study of the centralising Victorian administrative state.¹¹ British hegemony makes Britain the central site of a shift in the nature of regulation which rapidly globalises. Most importantly, its London and Irish models of a professionalised quasi-military police come to be utterly globalised. Peel is rightly seen as a pivotal figure, investing in police to increase the certainty of punishment while reducing its severity by eliminating capital punishment for most offences and replacing it with transportation. At the same time, his government embraces a debate on the future of another Benthamite move – the penitentiary. Beccaria and Bentham are justifiably seen as the central theorists of this third stage, though John Howard was the actor with practical influence in his own time.

Foucault seems mistaken to see the shift from the second to the third stages as *the* decisive modern penal shift. Garland shows for England that there is a late Victorian and Edwardian shift to an individualised, indeterminate, rehabilitative regime that rejects consistent Benthamite calibration of deterrence.¹² This is our fourth stage. In English prisons, this was more or less complete by World War I. It is part of a wider shift that gathers momentum throughout the first three quarters of the twentieth century – the rise of the welfare state with its therapeutic professions and the centralised Keynesian regulatory state. In some important ways the welfare state actually rises earlier in Bismarck's Germany and in colonies that are wealthier than Britain itself – New Zealand and Australia.¹³ The centralising regulatory project of the nation state that begins with the Passenger Acts ends with Keynes (and partly at the hands of Keynes).

Keynes is the most influential author of the Bretton Woods agreements. These render states as much objects as subjects of regulation at the hands of institutions such as the IMF, the World Bank and the GATT. At Havana in 1948 agreement is also reached for Keynes's international antitrust agency (the ITO). But national sovereignty fights back to pre-empt this regulator of states. Even so, as Chandler's work shows, the modelling of US antitrust regulation paradoxically fosters the global growth of massive multi-divisional corporations (as the alternative to

9 David Garland, *Punishment and Modern Society: A Study in Social Theory* (Chicago: University of Chicago Press, 1990) 140.

10 Foucault, n 2 above; Spierenburg, n 8 above; Garland, n 9 above, 158.

11 Oliver MacDonagh, *A Pattern of Government Growth, 1800–1860* (London: MacGibbon and Kee, 1961).

12 David Garland, *Punishment and Welfare: a History of Penal Strategies* (Aldershot: Gower, 1985).

13 Note the causal connection discussed later between transportation and the exceptionally early rise of the welfare state in Tasmania.

cartels).¹⁴ In 1995 for the first time, a majority of the largest 'economies' in the world are corporations rather than nations. States become objects as well as subjects of corporate regulation. Pre-eminently important to the regulation of states are the Big Five accounting firms and the corporations that give states their credit ratings – particularly Moody's and Standard and Poors. Like the stock exchanges and banks, the latter mediate and centralize the regulatory messages from the decentralised choices made in globalising markets. Reinsurers like Lloyds of London regulate what states can do in the regulation of national insurance markets. All of this has implications for how the regulation of property crime proceeds. For example, restorative justice for property crime becomes quite different in a world where insurers are the real bearers of most losses and policing is captured by an insurance actuarialism.¹⁵

The punitive/rehabilitative state is caught up in this fifth phase – which some refer to as the new regulatory state.¹⁶ Under the new regulatory state, there can be more state officials in business regulatory agencies than there are uniformed police for controlling individuals¹⁷ and there are more private than public police.¹⁸ At first the collapse of confidence in the state of the last quarter of the twentieth century, the 'nothing works' era in criminology, fuels a brief return to Beccarian classicism from 1975, particularly in the states which are at the 'cutting edge' of criminology – the US, the UK and Sweden. Neo-classicism quickly loses popularity among criminologists, however, as it is seen as rationalising rising Western imprisonment rates and as co-opted by law and order politicians. Obscure New Zealand emerges at the 'healing edge' of a new restorative justice, just as it is at the cutting edge of dismantling the Keynesian welfare state¹⁹. Singapore is another example of this conjuncture – at the same time as it is the most economically liberalizing state in the world (at least in free trade) it adopts 'restorative justice' as the mission of its juvenile justice system. This juxtaposition has a hard-edged fiscal side – restorative justice appeals as cheaper than the penitentiary for those who believe in the small state. It also fits with the new regulatory state idea that states should steer rather than row.²⁰ Restorative justice in the New Zealand model can be read as the state steering civil society, especially extended families, to take responsibility for delinquency control.

Clifford Shearing sees restorative justice and actuarialism in a late-Foucauldian frame of 'government at a distance'.²¹ Restorative justice in criminal law follows patterns set in the corporate sector. Decades before the restorative justice

14 Alfred D. Chandler, Jr, *The Visible Hand: The Managerial Revolution in American Business* (Cambridge, Mass: Belknap Press, 1977) and Alfred D. Chandler, Jr, *Scale and Scope: The Dynamics of Industrial Capitalism* (Cambridge, Mass: Belknap Press, 1990)

15 Malcolm Feeley and J. Simon, 'Actuarial Justice: The Emerging New Criminal Law' in D. Nelken (ed), *The Futures of Criminology* (London: Sage, 1994); David Garland "'Governmentality'" and the Problem of Crime: Foucault, Criminology, Sociology' (1997) 1 *Theoretical Criminology* 173–214.

16 G. Majone, 'The Rise of the Regulatory State in Europe' (1994) 17 *West European Politics* 77–101; M. Loughlin and C. Scott, 'The Regulatory State' in P. Dunlevy, I. Holliday and G. Peele (eds), *Developments in British Politics* 5 (London: Macmillan, 1997); Christine Parker, *Just Lawyers* (Oxford: Oxford University Press, 1999); John Braithwaite, 'The New Regulatory State and the Transformation of Criminology' (2000) 40 *British Journal of Criminology*, 222–38.

17 Peter Grabosky and John Braithwaite, *Of Manners Gentle: Enforcement Strategies of Australian Business Regulatory Agencies* (Melbourne: Oxford University Press, 1986)

18 Braithwaite, n 16 above

19 I acknowledge Howard Zehr as the author of the 'healing edge' metaphor in a restorative justice presentation I heard him give in Florida in 1998.

20 D. Osborne and T. Gaebler, *Reinventing Government* (New York: Addison-Wesley, 1992).

21 Clifford Shearing, 'Violence and the Changing Face of Governance: Privatization and its Implications' (Cape Town: Community Peace Foundation, 1997).

movement, the International Chamber of Commerce was globalising commercial arbitration as an alternative to state courts;²² courts operate in the shadow of informal commercial mediation. Consistent with Shearing's analysis, in Australia we were able to engage with praxis in what we now call restorative justice a decade earlier in corporate regulation than in criminal justice. While today restorative justice remains a radical alternative for the policing of serious common crime it is the mainstream for policing serious corporate crime.

The five stages of pre-state restorative justice, weak state corporal punishment, centralized Benthamite penitentiaries, Keynesian welfare state probation-prison-parole, and new regulatory state community policing are hardly neat. While I suspect they are more than an heuristic, fine-grained historical work is needed to flesh out their usefulness. The good thing about neat divides that have more than a grain of truth is that they are good to muddy. At all stages of human history, there is evidence of vibrant traditions of restorative justice surviving in civil society.²³ The aspiring kings of weak states and the Benthamite dreamers of strong states failed to kill it off. The prison was not born with Bentham's panopticon – the Greeks, Egyptians and Romans had them, as it would seem did most pre-modern civilisations.²⁴ Nor is there any period of history where torture of the body, executions and banishment have not survived somewhere in the West. This need not be inconsistent with a core Western historical sequence from restorative justice being more dominant to corporal/capital punishment being more dominant to imprisonment being more dominant.

My approach in this paper is indeed to muddy this five-stage core sequence by juxtaposing the history of a state that does not fit it very well (Australia) with one that fits it reasonably (the US). In the restorative justice movement there has been a lot of learning from 'Confucian' states like Japan, China and Singapore that deviate markedly from the posited core Western sequence. They have had such an ideological commitment to respecting robust extended families and communitarianism that we can study them to understand the possibilities for restorative justice in new Western regulatory states. For example, we have learnt from the empirical work of David Bayley and others on Japanese Koban policing and from Chinese *Bang Jiao* programmes.²⁵ And we have learnt from weak states like Papua New Guinea which have never acquired the strength to master either the restorative or the retributive traditions of their pre-state societies.²⁶

The method then is to formulate historical models of regulation that describe real changes, not just imagined changes, and then enrich our understanding of the limits of their truth-value by studying contexts where they are not so. This is the way I understand David Garland's project.²⁷ In this essay, the penal history of my own country is the foil for such a method.

22 Yves Dezalay and Garth Bryant, *Dealing in Virtue: International Commercial Arbitration and the Construction of a Transnational Legal Order* (Chicago: University of Chicago Press, 1996).

23 John Braithwaite, 'Restorative Justice: Assessing Optimistic and Pessimistic Accounts' in M. Tonry (ed), *25 Crime and Justice: A Review of Research* (Chicago: University of Chicago Press, 1999).

24 Edward M. Peters, 'Prison Before the Prison: The Ancient and Medieval Worlds' in Norval Morris and David J. Rothman (eds), *The Oxford History of the Prison. The Practice of Punishment in Western Society* (New York: Oxford University Press, 1995).

25 David H. Bayley, *Forces of Order: Police Behavior in Japan and the United States* (Berkeley: University of California Press, 1976), Hong Lu, 'Community Policing – Rhetoric or Reality? The Contemporary Chinese Community-Based Policing System in Shanghai' PhD dissertation, Arizona State University, 1998.

26 Sinclair Dinnen, 'Restorative Justice in Papua New Guinea' (1997) *25 International Journal of the Sociology of Law* 453–72.

27 n 9 above.

Two Eighteenth and Nineteenth Century Frontier Societies

The United States and Australia have much in common. They were both frontier societies during the same historical period – forged from rather similar surges of immigrant stock. Both frontiers were rich and suffered chronic labour shortages. Both were dominated by initial English immigration, had hugely influential Irish immigration during the nineteenth century, large influxes of Chinese during their gold rushes mid-century, and only later large surges of continental European immigration. Both advanced their frontiers through violence directed against the indigenous peoples of the land they occupied, indeed by genocide against some tribes. Influential historians of both nations have argued that their national identities were forged by the struggles at their frontiers.²⁸

Another thing they have in common is that at these European frontiers they forged the most influential penal ideas of the nineteenth century. This has always been recognised about US history; but it has been a suppressed fact of Australian history until recently. These US ideas were a failure in their own terms, though they were interpreted by Americans and Europeans as a success. The Australian ideas were a success in their own terms, though Australians, who believed them to be a success at the time, came to follow the English analysis (of Bentham) that they were a failure. De Tocqueville was the other European who was influential in defining the American penitentiary as a success and transportation of convicts to Australia a failure. Great distortion arose from both the shame Australia acquired about its past and the pride of Americans in theirs. Americans were fiercely proud of their republican penitentiaries in the 1820s and 1830s. Benjamin Rush juxtaposed them with 'Capital punishments ... the natural offspring of monarchical governments'.²⁹ At the very moment in American history when republican freedom was acquiring its deepest meaning, America took pride in institutions of unfreedom.³⁰ It became permanently attached to the myth that crime was a price of freedom,³¹ that freedom was so dangerous it had to be checked by remorseless unfreedom. The imprisonment rate in America has substantially and consistently increased since the mid-nineteenth century while it fell in Australia throughout the nineteenth century and beyond.³²

28 For the US see Frederick Jackson Turner, *The Frontier in American History* (New York: Holt, 1948), and for Australia see Russel Ward, *The Australian Legend* (Melbourne: Oxford University Press, 1958).

29 David J. Rothman, 'Perfecting the Prison: United States, 1789–1865' in Norval Morris and David J. Rothman (eds), *The Oxford History of the Prison: The Practice of Punishment in Western Society* (New York: Oxford University Press, 1995) 114.

30 David J. Rothman, *The Discovery of the Asylum: Social Order and Disorder in the New Republic* (Boston: Little Brown, 1971).

31 In referring to this as a myth, I am thinking, inter alia, of the evidence that societies with higher scores on political freedom have lower homicide rates (John Braithwaite and Valerie Braithwaite, 'The Effect of Income Inequality and Social Democracy On Homicide' (1980) 20 *British Journal of Criminology* 45–53).

32 M. Calahan, 'Trends in Incarceration in the United States Since 1880' (1979) 25 *Crime and Delinquency* 9–41, W. Clifford and R.W. Harding, 'Criminal Justice Processes and Perspectives in a Changing World' in *Australian Discussion Papers for the Seventh United Nations Congress on the Prevention of Crime and Treatment of Offenders* (Canberra: Australian Institute of Criminology, 1985) 50. Of course the imprisonment rate fell as a result of the cessation of transportation and the arrival of more free immigrants. However, it also fell dramatically long after transportation had ended, falling from 500 per 100,000 in the 1850s to 50 by 1920, see Satyanahu K. Mukherjee, John Walker and E. N. Jacobsen, *Crime and Punishment in the Colonies* (Sydney: University of New South Wales Printing Unit, 1985) 154.

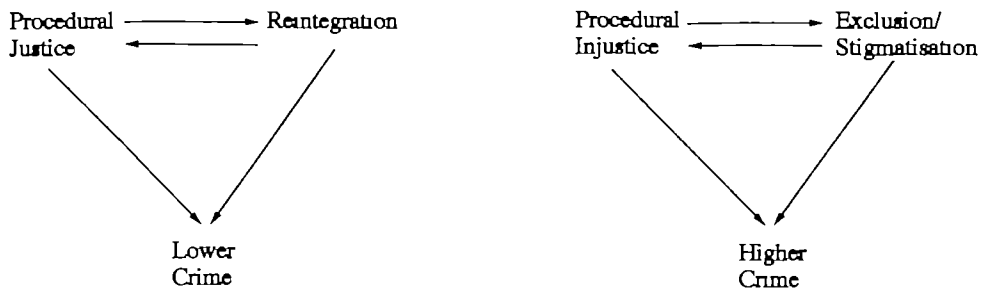
Australian convict society**US slave society**

Figure 1: Starting model for the analysis of a convict and slave society

I will argue that during the nineteenth century, Australia was transformed from being a high crime frontier society to a low crime society, while the US was transformed from a low to a high crime society. The key difference here is that to some degree the US in the 19th century is understood as a slave society, Australia as a convict society. This of course does not mean that most citizens in either case were convicts or slaves any more than to refer to frontier society implies that most citizens lived at the frontier. It means that the slave system, the convict system and the frontier had major and enduring effects on these societies. In the case of slavery, we might formulate this institution as important to understanding why well beyond the demise of slavery we can explain differences in crime rates in US cities by putting variables into regression analyses like percent black, black-white income differences and South/non-South (the Southern subculture of violence thesis).³³ Figure 1 summarises a theory of the way this difference played out.

These crucial differences between the two European frontiers were interconnected. Britain was shipping convicts to North America until the American Revolution, albeit in much smaller numbers than were later shipped to Australia. While American colonists initially wanted these convicts, by the time of the revolution African slavery was more profitable and capable of providing a more or less complete solution to its labour shortages. Selecting Africa's fittest workers made for higher labour productivity than having England ship its least desired workers. The continued vomiting of its unwanted criminals into North America was one of the colonists' many grievances against England. 'Send them back rattlesnakes', was Benjamin Franklin's rhetoric. George III vowed in a 1783 letter to Lord North that after the revolutionary war was won: 'Undoubtedly the Americans cannot expect nor ever will receive any favour from Me, but the permitting them to obtain Men unworthy to remain in this Island I shall certainly consent to'.³⁴ On the other side, after decades of unsuccessful attempts by the government of Virginia to get Britain to end the transportation of convicts, in 1740 it honoured a request from Britain for the Spanish war by sending them ex-

33 R.D. Gastil, 'Homicide and a Regional Culture of Violence' (1971) 36 *American Sociological Review* 412-27.

34 Manning Clark, *A History of Australia, Vol I: From the Earliest Times to the Age of Macquarie* (Melbourne: Melbourne University Press, 1962) 64.

convicts!³⁵ In 1775 Americans physically prevented convicts from landing, just as a shipload was sent home from Nova Scotia in 1789. It may be underestimated as a grievance motivating the Revolution, as might moral superiority to Britain be underestimated as an explanation for post-revolutionary pride in American penitentiaries

Australia became the receptacle for what were regarded as the dangerous classes of England and Ireland after the revolutionary war was lost. Indeed, it was colonised for the explicit purpose of acting as a penal colony. England never considered population of Australia by slaves; by then the English social movement against the slave trade was in full swing. In North Queensland, there was a brief and very modest experiment with Melanesian indentured labour on sugar plantations.

Up front, it is worth stating that in a nineteenth century society more transfixed by eugenics as the century proceeded, transportation supplies a natural experiment of the genetic explanation of crime. England in the nineteenth century hung its most dangerous felons and transported those it regarded as second only to the executed in levels of dangerousness. The convicts were what it regarded as its most degraded residuum. Tasmania had the highest number of convicts of any Australian state (42 per cent of the total), the most serious convicts were selected to go there rather than to the mainland, it had the lowest amount of free migration and experienced a continuation of transportation until 1853 (only Western Australia was later). Consequently, more than in any state, most of the genetic stock remained of convict descent until the end of the nineteenth century. By then Tasmania had a much lower crime rate than the rest of Australia. I will show it became one of the most serene places on earth by the 1880s. In contrast to recruitment from the high-crime British urban slums, US slave traders recruited from societies in which such ethnographic evidence as we have suggests stealing from fellow citizens was well regulated by restorative justice and banishment. The US slave traders recruited the fittest slaves they could find on the continent: those with the resilience of mind and body to survive the terrible voyage to America. They were the people the exporting society could least afford to lose.

The Kables Move to Sydney³⁶

Nineteen year old Susannah Holmes was sentenced to death in 1783 for breaking into a house and stealing a quantity of linen and silverware. The judge recommended that special circumstances warranted the extension of Royal Mercy, which was indeed extended. The sentence was commuted to transportation to America for 14 years. In 1786 while still awaiting transportation to a destination closed by the American Revolution, she fell pregnant in Norwich Castle Jail to another teenage housebreaker awaiting transportation, Henry Kable. Henry's father and older accomplice in the housebreaking had both gone to the scaffold; Henry had also won Royal Mercy. Susannah and Henry applied for permission to marry. This was refused.

In 1786 the first fleet of convicts to colonise Botany Bay was being assembled. When it was realised there was a shortage of female prisoners, the women in

35 A.G.L. Shaw, *Convicts and the Colonies* (London: Faber and Faber, 1966) 32.

36 I rely more than any other on David Neal's account of the Kables' story here, see David Neal, *The Rule of Law in a Penal Colony: Law and Power in Early New South Wales* (New York: Cambridge University Press, 1991) 1-9.

Norwich Castle Jail, including Susannah were transported to Plymouth. Desperately but unsuccessfully Henry pleaded to be allowed to marry and accompany Susannah and the baby. When Susannah arrived at the hulk *Dunkirk* in Plymouth harbour, the babe was torn from her on the captain's orders. John Simpson, a turnkey who observed this after ferrying Susannah out to the hulk, took the baby into his care, travelled with the infant to confront the unsympathetic staff of the Home Secretary Lord Sydney with the tragedy. Then he waited at Lord Sydney's house until he rushed up to him on his stairs to persuade him to allow Henry and Susannah to be married and reunited with their baby. Perhaps Lord Sydney was a politician with an eye for a good story, because Simpson's 700-mile round trip mercy dash with the child hit the headlines. A public subscription purchased a parcel of clothes and other valuables for the Kable family to take to Botany Bay.

In 1788 when the Kables disembarked at Botany Bay after an horrendous voyage below decks where many convicts perished, the parcel had disappeared. The Kables issued a writ naming the ship's captain as a defendant. In the first civil case in New South Wales, the court found for the convicts and ordered the captain to pay fifteen pounds to the Kables. It was a remarkable case because under English law at the time, the convicts had no right to sue. Even more remarkable was that they won their case against as powerful a figure as a ship's captain.

Just a couple of years later Henry became a constable of police and then for seven years Chief Constable of Sydney. He was dismissed from this position in 1802 after misbehaviour concerning conflicts with his trading activities. Chief Constable of colonial towns was a role many other ex-convicts subsequently filled, many corruptly. Henry and two partners subsequently became principal ship owners and among the very richest of the colony's mercantile class.

The Kables' story illustrates the central facts about transportation: the heartbreaking separation of families, the murderous voyage, the exploitation of convicts at the hands of the dregs of the British military who were often the bigger criminals, and hope, hope at the grace of a humane turnkey, but most significantly, hope from the fairness of a Governor and Judge Advocate who believed that unless convicts enjoyed the protection of the rule of law from the predations of their military jailers, a convict colony could not survive.

Procedural Justice in the Australian Convict Colony

A conclusion of this essay is that the Australian convicts and their children turned away from a life of crime. It is not my purpose, however, to argue that transportation was a good thing. It was not. There was the terrible suffering mentioned the last paragraph. But the more fundamental evil was that the crimes of the convicts were expiated through the theft of an entire continent from its rightful owners, mass murder of Aboriginal people, destruction and partial decimation of a splendid tapestry of cultures.

It is not my conclusion that the convict colony lacked brutality. In the use of the lash and the scaffold, it was horribly brutal. Historians differ on whether the lash was used more or less brutally in the Australian convict system than in the American slave system. I suspect that how one comes out on that question depends on which times and places one chooses to make the comparison between the two continents. But surely no place in America was as vicious as Captain Logan's Brisbane settlement, Captain Logan for whose murder both the Aborigines and the

convicts clamoured to claim credit.³⁷ On balance, I read the bits and pieces of records on the frequency and severity of flogging as showing that the flogging of Australian convicts probably was more brutal than that of North American slaves. It was not restricted to men until 1817, the number of lashes was awesome, at its worst it spattered the floggers' faces with blood leaving pieces of flesh scattered over the ground.³⁸

More brutal though it may have been, Australian flogging was ordered in a more procedurally fair fashion. While American slave-owners stood over their property with a whip, administering it on the spot, Australian masters of convicts had to send them to a hearing before a magistrate before the lash could be administered by a constable. Many of the constables were ex-convicts who might be bribed to be less brutal with their work³⁹. This was institutionalised at first as part of Governor King's programme to regulate the abuses of power of the New South Wales Corps.⁴⁰ Consider in historical context the procedural innovation involved here. Masters of convicts were being required to have their corporal punishment authorised by a court when British naval and military commanders were not so constrained, when masters under English common law could flog apprentices and indentured workers on the spot, schoolmasters could do so to students, and it was not long since husbands had a right/duty to do so to recalcitrant wives.

As illustrated by the Kables' case, Australian convicts had a right to hold property and sue to protect it, to sell part of their labour, to appear as witnesses in court cases, and to write petitions to a Governor who mostly treated them seriously. English prisoners did not enjoy these rights. Convicts could and did press charges against their masters for ill-treatment in ways that are impossible in contemporary Australian prisons.⁴¹ Convicts were well advised to be sure of their ground before complaining. Atkinson reports that there were 210 charges against convicts by masters before the Scone bench in 1833 compared to only six by convicts against masters.⁴² But all six did succeed. While magistrates' courts administered a rough justice even worse than contemporary lower courts, between a quarter and a third of those prosecuted before the colony's criminal court were acquitted.⁴³ Perhaps most remarkably, convicts assigned to work for landowners could obtain a writ of *habeus corpus* to protect them from being locked up without trial. Without a court order a convict could not even be put in irons for any reason other than prevention of escape. In a famous case Justice Stephen in 1827 upheld a writ of *habeus corpus* from some convicts who had been locked up for five or six weeks for cattle stealing without being sent to court. The judge ordered the prisoners to be released, finding that "the rights of prisoners were as sacred in the eye of the law as those of free men".⁴⁴ Contemporary criminal lawyers would view this as an extension of *habeus corpus* to prisoners under sentence to contest an administrative decision to

37 Raymond Evans and William Thorpe, 'Power, Punishment and Penal Labour: Convict Workers and Moreton Bay' (1992) 25 *Australian Historical Studies* 90-111.

38 n 36 above, 36-37; and especially at Moreton Bay, Evans and Thorpe, *ibid.* On the other side, see Stephen Nicholas, 'The Care and Feeding of Convicts' in S. Nicholas (ed), *Convict Workers. Reinterpreting Australia's Past* (Sydney: Cambridge University Press, 1988) 180-183; J.B. Hirst, *Convict Society and Its Enemies: A History of Early New South Wales* (Sydney: George Allen & Unwin, 1983).

39 Hirst, *ibid* 61.

40 *ibid* 58.

41 *ibid* 109-111.

42 Alan Atkinson, 'Four Patterns of Convict Protest' (1979) 37 *Labour History* 28-51.

43 Hirst, n 38 above, 113.

44 *ibid* 118.

reclassify them from a work release programme to maximum security following a further criminal offence. A courageous judgement; yet Chief Justice Forbes backed Justice Stephen when he was subjected to some political pressure over it. It was not an isolated one. A year after the English courts ruled that questions about previous offences could not be asked during criminal trials, Judge Willis refused to allow the NSW Attorney-General to ask a witness 'what were you sent out for?'.⁴⁵ Hence, Neal concludes that the courts acted as a *de facto* parliament:

The American and French revolutions gave political actors in New South Wales recent models for political change. Neither the ideology of universal rights nor the strategy of armed revolution was adopted in New South Wales. The presence of Jacobins, Irish rebels and political leaders who were well versed in those ideas and strategies meant that the strategies actually adopted were not adopted in ignorance of other possibilities. [Instead] protagonists relied on their British birthrights and deployed the language of the rule of law to secure them and to forge new social and political order out of the penal colony at Botany Bay.⁴⁶

After the abortive Irish Rebellion of 1804, there was no convict uprising. A single settler employing a number (sometimes dozens) of convicts hundreds of miles from the reach of state authority would seem to have reason for fear. Yet the Australian bourgeoisie lived less in fear of a rebellion of their dangerous classes than was manifest in the fears of the European bourgeoisie of the era or in the fears of American plantation owners of slave uprisings. That, I will argue, was because it was clear to them that the convicts had hope, a stake in the future,⁴⁷ and some prospect of fair procedure to deal with the injustices of the present. The literature of the social psychology of procedural justice shows that even in the context of a harsh criminal justice system such as that of the contemporary US, adverse outcomes combined with a perception of fairness of procedures can deliver high compliance with the law.⁴⁸ One reason is that when one shares an identity as a citizen of a just legal order, there is a willingness to comply with that order.⁴⁹ To realign the identities of convicts to those of law abiding citizens, convicts need to be persuaded that they are now in reach of a society where the rule of law is something that offers practical protection to them and is therefore worthy of being honoured. Brutality is more bearable when its end can be imagined and seen and when its excesses can be challenged by fair procedure. Neither Australian Aborigines nor American slaves could imagine its end in the same way the white Australian convicts could. It is the combination of adversity with perceived injustice that conduces toward defiance of a legal order.⁵⁰

While the adversity of floggings was terrible, mostly it could be avoided by sticking to the rules. The convicts were after all in an utterly open prison where 'the lash had to do the work of the walls, the warders and the punishment cells'.⁵¹

45 *ibid* 119.

46 n 36 above, 25.

47 '[E]ach governor, at least until the end of Macquarie's term of office, was diligent in engendering belief in ... providing convicts and emancipists with "a something to lose"' (W. Nichol, 'Ideology and the Convict System in New South Wales, 1788-1820' (1986) 22 *Historical Studies* 13).

48 E. Allan Lind and Tom R. Tyler, *The Social Psychology of Procedural Justice* (New York: Plenum Press, 1988); Tom Tyler, *Why People Obey the Law* (New Haven, Ct.: Yale University Press, 1990).

49 Tom Tyler and Robyn M. Dawes, 'Fairness in Groups: Comparing the Self-Interest and Social Identity Perspectives' in B. A. Meesters and J. Baron (eds), *Psychological Perspectives on Justice: Theory and Applications* (Cambridge: Cambridge University Press, 1993).

50 Lawrence W. Sherman, 'Defiance, Deterrence and Irrelevance: A Theory of the Criminal Sanction' (1993) 30 *Journal of Research in Crime and Delinquency* 445-73.

51 *Hirst*, n 38 above, 68-69.

Moreover, the leading histories agree that 'the common wisdom remained that kindness and firmness rather than harshness was the best method of dealing with them'.⁵² New South Wales even had an early social scientific researcher who interviewed convicts as to whether they would behave better for a good master or a bad one. Many found the question awfully funny, but those who were persuaded to take it seriously felt they would behave better for a good master.⁵³ Flogging aside, in many respects the adversity of male convicts was less than that of employed men in England. They were better fed than workers back in England. Indeed they received dietary intakes recommended by contemporary Australian standards for younger men and more than Scottish coal miners received in the 1980s.⁵⁴ The convicts worked shorter hours,⁵⁵ were better housed,⁵⁶ better clothed,⁵⁷ and had better access to medical care⁵⁸ than both American slaves and free English workers (but see the questioning of this conclusion for Moreton Bay by Evans and Thorpe).⁵⁹

How can we make sense of this degree of good treatment? How can we make sense of some of the excess assurance that procedural rights allowed convicts at times, as evidenced by the following:

'Do it yourself and be buggered; punish me and be damned', was how one convict greeted his master's enquiry as to why he hadn't done his work. When a magistrate sentenced a young woman to ten days' solitary confinement on the complaint of her master [women could not be flogged after 1817], she turned and spat in her master's face. The magistrate then increased the sentence to 30 days which was the maximum, as the young woman well knew. 'O, thank you, I am much obliged to your worship; thirty days; I am very fond of an odd number, would you be kind enough to indulge me, and make it thirty-one days; do, your worship, I should like to have thirty one'.⁶⁰

The answer to why this was possible is that development of the colony was held back by labour shortages. First, early Governors of the colony formed the view that their primary objective was to get their convict population to be willing workers and that the best way to accomplish this was to institutionalise just treatment. We will see that early implementation of the policy achieved early success after a difficult first few years where chronic starvation confronted the settlement. Second, the judiciary played a central role in pushing procedural justice for convicts in Australia well beyond that enjoyed in England or America. Third, most convicts were put to work by assigning them to masters who believed that they could get the best out of them by being fair but firm. There were brutal masters, but these did not grow rich in the way the fair and firm masters did. A brutal master like James Wright in Canberra (at Lanyon) could turn no fewer than eight of his convicts into bushrangers and thereby render himself bankrupt. Collective agreements among the convicts to work slow for masters they did not like so that the master would send them back to be reassigned to a better master were effective weapons of the weak. Convicts could sometimes pick their own master by conspiring with a wealthy landowner who was a good master to support them in a

⁵² *ibid* 71.

⁵³ *ibid* 73.

⁵⁴ Nicholas, n 38 above, 185

⁵⁵ *ibid* 187–88.

⁵⁶ *ibid* 189–91.

⁵⁷ *ibid* 194–95.

⁵⁸ *ibid* 192–94.

⁵⁹ n 37 above.

⁶⁰ Hirst, n 38 above, 70–71.

complaint to the governor or a magistrate (who might be a friend of the wealthy landowner) against a bad master for ill-treatment. When the Governor or a Court took convicts off the unjust master, the wealthy landowner could grab them. Or the convict could simply punch the Master on the nose and cop a flogging followed by reassignment.

With that as a main conclusion, there is a need to qualify it considerably. All of the early Governors enjoyed largely unchecked local power and all abused it most unjustly on issues that were important enough to them. Indeed, this was especially true of the greatest, the most visionary and the most reintegrative of the early Governors, Lachlan Macquarie.⁶¹ One of the less great Governors was Captain Bligh, of the *Bounty* fame. His most remembered contribution to the procedural justice debate was: '[t]he law sir! Damn the Law; my will is the law, and woe unto the man that dares to disobey it!'.⁶² Bligh's Governorship fell to the only military coup in Australian history. While he was a more complex character than one might gather from the portrayals of him by Errol Flynn and Mel Gibson,⁶³ he was a man who might have done better in life had he read Tom Tyler's research on the social psychology of procedural justice.

The brutality of Captain Logan at Moreton Bay was not matched by all of the military commanders of the outlying penal establishments where convicts were sent for punishment or of the Women's Factory in Parramatta where many female convicts were incarcerated. Yet most of the commanders of the Norfolk Island prison were rivals to Logan in brutality and injustice. While assignment to work that involved freedom to roam the countryside was the fate of the majority, we must not forget that there was a diversity of convict experiences of Australia. Many did die in chains and in conditions of relentless, savage confinement. More still at the scaffold and on the voyage. Yet there is a need to balance the image of the convict era which tends to focus exclusively on its wretchedness and hopelessness. The tour guide at Port Arthur points to a cliff and says that is where the two convict boys leaped to their death as described in Marcus Clark's novel, *For the Term of His Natural Life*. The boys were at Point Puer, an institution for juvenile offenders established four years before the first such institution was established in England.⁶⁴

'I can do it now,' said Tommy. 'I feel strong.'

'Will it hurt much, Tommy?' said Billy, who was not so courageous.

'Not so much as a whipping.'

'I'm afraid! Oh, Tom, it's so deep! Don't leave me, Tom!'

The bigger baby took his little handkerchief from his neck, and with it bound his left hand to his companion's right.

'Now I can't leave you.'

'What was it the Lady that kidded us' said, Tommy?'

'Lord have pity on them two fatherless children!' repeated Tommy.

'Let's say it, Tom.'

And so the two babies knelt down on the brink of the cliff, and raising the bound hands together, looked up at the sky, and said, 'Lord have pity on us two fatherless children!' And then they kissed each other, and did it.

61 John Ritchie, *Lachlan Macquarie* (Melbourne: Melbourne University Press, 1986).

62 Bruce Kercher, *Debt, Seduction and Other Disasters: The Birth of Civil Law in Convict* (New South Wales, Sydney: Federation Press, 1996) 10.

63 Greg Denning, *Mr. Bligh's Bad Language: Passion, Power and Theatre on the Bounty* (Melbourne: Cambridge University Press, 1992).

64 Shaw, n 35 above, 244.

While Robert Hughes says that no incident like this is known to have occurred at the boy convicts' settlement at Point Puer,⁶⁵ we know that suicide was not uncommon among the convicts. While most of the boys at Point Puer were confined there only for long enough to teach them a trade, while many of them went on to become the business elite of Hobart Town, many were also subjected to remorseless brutality. Simplifying, we might say there are two basic stories of the convict experience. The majority story is one of assignment to work in the Australian bush or as a servant in town (especially for women) by masters who were fair to those who worked well. Yet those same masters could be vicious with defiant convicts. A tiny minority of masters were vicious with all of them. Defiant convicts who were treated savagely tended to become more defiant (see, for example, the analysis of Judge Therry of bushrangers as men who had been repeatedly and unjustly flogged).⁶⁶ This put them on a career trajectory to Norfolk Island or some other hell, to the scaffold or to life as a bushranger. The legends of Botany Bay are brimming with stories – here is a fictional and factual one of such defiance, both from the pen of Charles Macalister:

And some dark night when everything is silent in the town
I'll kill the tyrants, one and all; and shoot th' Floggers down:

I'll give th' law a little shock: remember what I say,
They'll yet regret they sent Jim Jones in chains to Botany Bay.

Charles Macalister, *Old Pioneering Days in the Sunny South*

[They] were made of the sternest human stuff possible, and men of that type never flinched under the lash. On two occasions I saw men – after undergoing, one a flogging of fifty, and the other seventy-five lashes, bleeding as they were, deliberately spit, after the punishment, in the flogger's face. One of them told Black Francis 'he couldn't flog hard enough to kill a butterfly'.⁶⁷

In the next section, we tell the majority story, which is one of reintegration into respectable society, of the reformed criminal. Yet we should not forget the minority who were made worse by their convict experience (Figure 2 summarizes the two trajectories). Certainly, the reformed majority did not. They sang the ballads that lionized the likes of Bold Jack Donohoe, they protected the bushrangers from the police. This was because they had known their own Jack Donohoe; they had seen how the system that was forgiving to them could be so vengeful to others. Remarkably, to some degree this sentiment survives in Australia at the end of the 20th century. I noticed it on my son's primary school graduation night, when the children performed the play 'Ned Kelly'. In what other country would this be thought appropriate to such an occasion: a play that treats lovingly a recidivist armed robber who callously murdered policemen. It is part of the Australian legend to feel sorry for Ned when he hangs, to agree with him when he tells the judge he will meet him in hell, to resent the police, to be at least somewhat on his side because of the way they treated his Irish convict mum. Even the policeman-parent who sat near me did not think of complaining about the play. You might say Americans have their Jesse James or Billy the Kid, as do other countries. But Ned Kelly is nearer to Jefferson or Lincoln in the sense that matters. For twentieth century Australians, he is the figure from their nineteenth century

65 Robert Hughes, *The Fatal Shore: The Epic of Australia's Founding* (New York: Alfred A. Knopf, 1987) 602.

66 Judge Therry quoted in Ward, n 28 above, 138.

67 Charles Macalister quoted in Ward, n 28 above, 34.

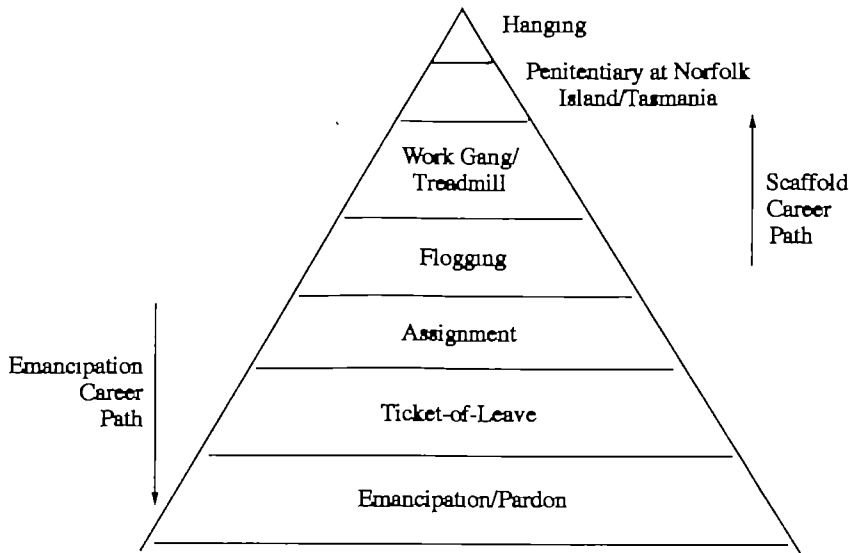


Figure 2: Two Career Paths Out of the Enforcement Pyramid of Australian Convict Society

history who is most widely known. Perhaps it is perverse that more Australians have a soft spot for Ned than for Macquarie or Parkes or Macarthur, but this is the fact of the matter.

So we have simplified the procedural justice-injustice story into one of an assigned majority who went straight and a brutalised minority who responded to injustice with escalated defiance. We must at least complicate this simplification now with a crucial third category. Henry Kable falls into it. These were ex-convicts, emancipists, who were reformed from being a powerless underclass criminal in England to being a respectable citizen of the colony who prospered through crimes of the powerful. Such men modelled their exploitative respectability on the illegal trading activities of the officers of the New South Wales Corps. This started in the 1790s when the officers used their power to secure trade monopolies, forbidding anyone else boarding newly arriving ships and buying up all the stock so they could sell it at extravagant mark-ups.⁶⁸ Many took their fraudulent business practices into the corrupt conduct of high government office. The tradition continued in the twentieth century: Allan Bond, convicted for burglary as a young English immigrant became a business icon and a competitor for title of the biggest corporate criminal any nation's history has known (in terms of the sheer dollar amounts misappropriated).

While no nation may be able to match Australia's record of transforming powerless criminals into productive law-abiding citizens, nor perhaps can any country match its record in turning them into criminal abusers of power. The worst abuses involved mass killings of Aborigines to take vast tracts of land from them. A common career path for the most ruthless 'reformed' criminals in New South Wales was the police. Police were neither well paid compared to hard workers on

⁶⁸ Clark, n 34 above, 135.

the land, nor were they well liked. Yet they could make money by being corrupt. This was a configuration of incentives that tended to attract the dregs of convict society into policing. The emancipists themselves regarded other ex-convicts who would choose to earn a living that involved flogging convicts as below respect.

A remarkable statistic for 1825, when the average strength of the Sydney police force was 50: 57 officers were dismissed for misconduct and 25 resigned in one year.⁶⁹ An equally stunning statistic from Sturma's study of convictions for serious crimes before the Superior Courts of New South Wales is that the occupation of 10 per cent of those convicted in 1841 was 'policeman'.⁷⁰ Things improved, as O'Malley has argued, with this percentage dropping to one per cent a decade later.⁷¹ Most of the early constables and chief constables were, like Henry Kable, former convicts. Many, perhaps most of those who rose to the top were corrupt. As late as 1844 in Melbourne (after transportation to the Eastern mainland had ceased), Chief Constable William Sugden articulated a police policy to enlist emancipists as detectives 'because they were better acquainted with the nature and character of arrivals from Van Dieman's Land'.⁷² Things did not improve much in New South Wales until late in the century,⁷³ even after the *Sydney Police Act, 1833* sought to regularise policing on the model of the London Metropolitan Police. The first police commissioner under the new Act was dismissed over corruption charges and the three succeeding office holders were similarly dismissed for alleged improprieties.⁷⁴ These were not always emancipists. Among the three was William Augustus Miles, supposedly the illegitimate son of King William IV appointed through Royal patronage. This was the other side of the white-collar crime problem in Australia: as well as being a dumping ground for convicts, it was also a dumping ground for ne'er do wells who the British ruling classes wanted to keep at a distance. These the most notionally respectable of the colonials (Wentworth being the most influential character of the type) were role models for the most ruthless emancipists. While abuse of power was attenuated when the Australian colonies acquired more robust democratic constitutions with proper separations of powers, there had nevertheless been an embedding of a culture of abuse of power in certain institutions. Hence, for example, in the twentieth century the New South Wales police managed to repeat the feat of a sequence of three corrupt commissioners.

Reintegration

Assignment was the principal vehicle of reintegration. There were cases where wives of convicts became free immigrants to Australia, set up small businesses and succeeded in having their husbands assigned to work for them. More generally marriage was encouraged in the colony because stable family relationships were seen as vital to the rehabilitation of criminals. Men confined in the Sydney convict

69 Pat O'Malley, *Law, Capitalism and Democracy: A Sociology of Australian Legal Order* (Sydney: George Allen & Unwin, 1983) 53.

70 Michael Sturma, *Vice in a Vicious Society. Crime and Convicts in Mid-Nineteenth Century New South Wales* (St. Lucia: University of Queensland Press, 1983) 88.

71 n 69 above.

72 Robert Haldane, *The People's Force: A History of the Victoria Police* (Melbourne: Melbourne University Press, 1986) 17. See John Blackler, 'The Case of the Missing Australian Policeman: the Investigator in the Fiction of the "Continent of Mystery"', unpublished manuscript, 1998.

73 n 69 above.

74 Blackler, n 72 above.

barracks were allowed to sleep with their wives in the town and were given Friday and Saturday to themselves so they had better prospects of supporting their families. Well-behaved convicts with wives in Britain were eligible to have them and their children shipped to the colony free of charge (from 1816).⁷⁵

The economic reintegration policies were even more humane and innovative than the family reintegration policies. Most convicts could expect to be assigned to work for a master. Mostly this meant a private landowner, but for many it also meant working as domestic servants in towns or for government. Convicts were paid modestly at government regulated rates for this work. But in conditions of acute labour shortage, many masters secured the best convicts by tempting them with side-payments. In addition, once convicts completed their assigned work or assigned hours, they had a right to work at market rates for other masters. As market wage rates were high in conditions of labour shortage, many convicts made good money while on assignment. Responsible conduct on assignment would entitle a convict to a 'ticket-of-leave'. With a ticket-of-leave convicts could work for whoever they chose. The idea of the ticket-of-leave was copied in England later in the century and became the institution that today we call parole. Note that the desire of masters to hang on to their best convicts once they were entitled to a ticket-of-leave was another reason for treating them fairly while on assignment. Finally, pardons were widely granted by the government to well-behaved, highly productive convicts, especially by Governor Macquarie.

We might even say that the term 'restorative justice' was invented by Lachlan Macquarie early in the nineteenth century. He often spoke of 'restoring' convicts to their place in respectable society and was regularly denounced for this usage by exclusives (free settlers) who pointed out that convicts were not members of respectable society at the time of their conviction and Macquarie was in fact elevating them above their station. Another New South Wales invention was the 'certificate of freedom' issued to convicts who were pardoned or who had served their terms. It was necessary in a convict society where sometimes emancipists needed to prove they were free. The certificate claimed that the holder was 'restored to all the rights and privileges of free subjects'.⁷⁶ This was another example of where colonial law bestowed more full rights on ex-convicts than English law of the time, where even a pardon was not sufficient to restore certain rights to an ex-offender.

Emancipated convicts were given substantial free grants of land, animals, tools and seeds, sufficient for them to become economically viable settlers. Many became large landholders. They became masters of convicts on assignment themselves. By the 1820s, the fourth decade after the first convicts arrived, ex-convicts were masters of the majority of convicts on assignment, owned over half the wealth of the colony and three quarters of the land.⁷⁷ Here is where the contrast with economic integration for American slave society is sharpest. Imagine how different American society would be today if by that date half the wealth in the South was owned by former slaves, if slaves had been given slaves to become slave-owners themselves after just a few years in chains.

The emancipists grabbed their opportunities with both hands. James Ruse was the first emancipist to be granted land in 1790. He was an experienced farmer in

⁷⁵ Hirst, n 38 above, 80.

⁷⁶ *ibid* 108.

⁷⁷ *ibid* 81.

England and worked out what the military farmers had not managed to do – how to cultivate the soil of the great south land to make it productive. Ruse also introduced the first of many profit-sharing schemes with the convicts assigned to work for him. Emancipists were mostly fair, even generous, in their treatment of convicts under sentence. In a scarce labour market this was another factor that put pressure on the exclusives to hold their convicts by also being fair.

Not only did economic integration thus contribute to procedural justice as a result of this scarce labour market, procedural justice also contributed to economic reintegration. An engineer brought out from England to employ 100 convicts digging a tunnel to carry water, John Busby, kept the best records on how he managed his workforce. At first he achieved good results by sending convicts who behaved badly during the week to work at the colony treadmill on Saturday. After adverse comment on this practice by a magistrate, Busby was advised it was an illegal punishment unless a magistrate approved it. Rather than put up with the delays of sending convicts to court, Busby set them a task for the day which, once completed, allowed the men to be free to earn extra income elsewhere at market rates.⁷⁸

Perhaps more surprising than the integration of emancipists into agriculture and industry was their integration into the professions. William Redfern was appointed principal surgeon, the first of a number of emancipists to rise to the top of the profession of medicine. The first practicing lawyers in the colony were convicts on tickets-of-leave. Macquarie appointed three non-lawyer emancipists as magistrates. Obviously, this would not have been permitted in England. Macquarie's pre-eminent government architect, Francis Greenway, still the most famous architect in Australian history, was a convict. Education was only allowed to be given priority in the colony early in the nineteenth century by appointing convict teachers on tickets-of-leave.

Perhaps the most strategic integration of emancipists into the colony was that from early in the history of the colony emancipists ran successful newspapers. The emancipist press gave Governors a hard time when they buckled to the demands of exclusives.⁷⁹ Its very existence consolidated the legal and economic gains emancipists had secured.

In addition to family and economic reintegration, the convicts enjoyed a surprising level of social reintegration. Macquarie was the first Governor to invite ex-convicts to his table, a gesture exclusives resented. Exclusives did respect a strong norm against referring to a convict as a convict, however. The most usual respectful appellation was 'government man'. In all but a few special places or circumstances convicts under sentence wore dress that was varied; in appearance they were indistinguishable from other citizens. While hurtful stigmatisation did occur, there was a lot less of it than convicts suffered in England or perhaps anywhere else.

More important to Macquarie than family, economic and social reintegration was reintegration into the church. To this end, welcoming convicts into the bosom of the church was attempted by promoting the work of the British and Foreign Bible Society and the Sunday School movement.

Unless crime was of such a serious kind that it could not be kept from the police, crimes of one emancipist against another were mostly dealt with informally, if not

⁷⁸ *ibid* 66.

⁷⁹ For instance, see Manning Clark, *A History of Australia, Vol II: New South Wales and Van Diemen's Land 1922–1838* (Melbourne: Melbourne University Press, 1968) 74–75

always justly by 'kangaroo courts'. Ward's description of how 'collective public opinion' rather than 'state power' maintained order on the goldfields is instructive, as is his observation, astute observer of bush institutions as he was, that 'roll-ups' at Western Australian goldfields and Northern Territory mines were still in the 1930s the method for allowing all miners to participate in the adjudication of justice.⁸⁰

These then are the reasons for my interpretation that Australian convict society experienced high levels of reintegration that promoted high levels of procedural justice and vice versa. While I have not documented the story of the opposite claim about US slave society, I take it to be uncontroversial that slave society was characterized by stigmatisation rather than reintegration, stigmatisation that was mutually constitutive of high levels of procedural injustice. But did the procedural justice and reintegration – the policy of restorative justice – result in lower crime rates, as the theories of restorative justice,⁸¹ reintegrative shaming⁸² and procedural justice would predict?⁸³

Restorative Justice and the Crime Rate

[I]t is the first time anyone has dared to fashion a society from all that is wicked in another⁸⁴

Jeremy Bentham saw transportation as a threat to his panopticon. He branded it an abject failure as a crime prevention strategy. This was the view that came to prevail in England; it influenced that intellectual of the other great power of the period, de Tocqueville. Australian thinking mid-century came to follow the European orthodoxy. Yet during the period of transportation, Australian thinking was overwhelmingly that transportation worked in reforming criminals. Typical was the view of longstanding New South Wales Attorney-General John Plunkett in 1840 that the convict system: 'has reformed more than any penitentiary or any other system of punishment that has hitherto been discovered'.⁸⁵ During the transportation era this was the dominant analysis in the British bureaucracy as well. But British party politics rendered a different story. The Tory side tended to buy Bentham's view that transportation provided uncertain and therefore weak deterrence. They seized on stories of ex-convicts making their fortunes and questioned what kind of deterrence of crime this was. Both the Whigs and Radicals came to equate the convict system with slavery; Wilberforce felt it was degrading for Christian farmers to be masters of slaves. They seized upon stories of flogging and depravity in the colony. Both sides alighted upon travellers' tales that sodomy was rife among male convicts and that female convicts were widely used as whores. This was also a major theme of Tocqueville's critique: 'What Australian society essentially lacks is morals. And how could it be otherwise? ... [Women] lost those traditions of modesty and virtue which characterise their sex in the

80 Ward, n 28 above, 110–111.

81 n 23 above.

82 John Braithwaite, *Crime, Shame and Reintegration* (Melbourne: Cambridge University Press, 1989).

83 Tyler, n 48 above.

84 Citizen Toulangeon, 3 January 1803, quoted in Colin Forster, 'Part I: Prison and Deportation: Botany Bay and the French Penal Debate, 1814–32', *Working Paper in Economic History*, No. 136 (Canberra: Australian National University, 1990).

85 John N. Molony, *An Architect of Freedom: John Hubert Plunkett in New South Wales, 1832–1869* (Canberra: Australian National University Press, 1973) 121.

mother country ... bastards still make up a quarter of the children'.⁸⁶ Like Tocqueville, Commissioner Bigge, who was sent out to reign in Governor Macquarie's excesses of integration, saw a need 'to preserve British 'civilisation' from the levelling tendencies of frontier society'.⁸⁷

This was an unassailable Whig-Tory political cocktail, combined with the intellectually respected analyses of Bentham and Tocqueville, for redefining transportation as a policy failure. Yet Australian and English bureaucratic opinion held out against these odds for many decades. Both Bentham in England and Tocqueville in France lost the debates during their own lifetimes, never to see their beloved penitentiaries built. In the twentieth century, the serious historians of the Australian convict experience came to redefine it again.

Robert Hughes concludes with the sweeping claim that 'the assignment system in Australia was by far the most successful form of penal rehabilitation that had ever been tried in English, American or European history'.⁸⁸ Australia's leading historians of the convict era mostly agree that it was a violent, uncivil, drunken, corrupt place at first which became more civilized as most convicts turned away from crime to become productive citizens and bore children who became remarkably law abiding. I am thinking here particularly of the histories of Manning Clark, Keith Hancock, John Hirst, Robert Hughes, A.G.L. Shaw and L.L. Robson.⁸⁹ But I am also thinking of the first scholarly history of Australia, written by Ernest de Blosseville in 1831, revised in a second edition in 1859, and recently made accessible to Australian historians by Colin Forster.⁹⁰ Blosseville's book attracted great interest in a France that was fascinated by the idea of Australia; it won the prestigious Montyon prize of the Academie Française in 1832, a prize that de Beaumont and de Tocqueville won the next year on the opposite side of the debate for their book advocating the American penitentiary.⁹¹

Blosseville believed that after only four years the colony had been able to put an early disorderly period behind it, that crime had diminished, that the streets of Sydney were by then safer than some of the streets of London. He attributed particular effectiveness to pardons, economic reintegration of offenders and selecting convicts of good character for the Sydney night watch.⁹² Blosseville relied on the interesting resource of a French government fact-finding mission which spent five months in New South Wales in 1802. Péron reported for the members of the mission:

Never perhaps has a more worthy object of study been presented to a statesman or philosopher ... There, brought together, are those terrible ruffians who were for so long the terror of the government of their country: thrust from the bosom of European society ... The majority, having atoned for their crimes by a hard bondage, have rejoined the ranks of the citizens. Obligated to concern themselves with the maintenance of law and order to safeguard

86 Quoted in Colin Forster, *France and Botany Bay: the Lure of a Penal Colony* (Melbourne: Melbourne University Press, 1996) 104.

87 John Ritchie, *Punishment and Profit* (Melbourne: Heinemann, 1970) 220.

88 n 65 above, 103.

89 Manning Clark, *Manning Clark's History of Australia*, abridged by Michael Cathcart (Ringwood: Penguin, 1992); Sir Keith Hancock, *Australia* (London: Ernest Benn, 1930); n 65 above; n 35 above; L. L. Robson, *The Convict Settlers of Australia: An Enquiry into the Origin and Character of the Convicts Transported to New South Wales and Van Diemen's Land* (Melbourne: Melbourne University Press, 1965).

90 n 86 above.

91 Gustave de Beaumont and Alexis de Tocqueville, *On the Penitentiary System in the United States and Its Application in France* (Carbondale and Edwardsville: Southern Illinois University Press, 1964)

92 Forster, n 86 above, 79

the property they have acquired, having become nearly at the same time husbands and fathers, they are bound to their present state by the most powerful and beloved ties. The same revolution, brought about by the same means, has taken place in the women; and miserable prostitutes, gradually restored to more proper principles of conduct, are today bright and hard-working mothers of families.⁹³

Arago, who had been a draftsman on the 1802 expedition, revisited Sydney in 1819, concluding in extravagant terms that there had been further progress:

[T]here the forger is employed in useful work, which at first gives him land, then esteem, and finally honours. There the thief, abjuring his blameworthy habits, often attains the magistracy, and even becomes here the scourge of thieves. I have seen a swindler, now honoured with the just confidence of the Government, bestowing upon the children of Sidney as much by his example as by his teaching, principles of the strictest virtue and the greatest honour. One would say that the air of this country, though savages breathe it, purifies the mind and makes every noble sentiment grow within it.⁹⁴

Further French writing on the virtues of the Australian penal colony was prompted by the Maçon Society essay competition of 1827 which invited essayists to attempt a republican analysis of punishment:

Outline, for the replacement of hard labour, a punishment which, without ceasing to satisfy the needs of justice, leaves less degradation in the soul of the condemned; propose measures to take in the meantime so that freed *forçats* are no longer driven to misery by a public opinion which rejects them, and so that their presence no longer threatens the society which receives them.⁹⁵

Tocqueville was the formidable figure on the other side of the debate. Tocqueville, like Bentham, believed transportation created a society 'composed of vicious elements which sooner or later form a people difficult to govern and dangerous to free'.⁹⁶ Blosseville was a good friend of Tocqueville and well aware of his views. His conclusion was that Tocqueville had got it wrong, a conclusion based upon Blosseville's careful empirical engagement with what was actually known about the workings of the English penal colony.

I too suspect that it was Blosseville and the contemporary Australian historians who engaged more carefully with such evidence as we have than did Bentham and Tocqueville. Admittedly, it is not evidence of a quality that will be definitive to any contemporary criminologist, including this one. Before turning to that evidence, however, we must consider a matter on which there is raging disagreement among the Australian historians: were the convicts part of a criminal class?

Were the Convicts Serious Criminals to Start With?

From 1913 a reaction to the English reading of the convict experience set in. Watson and Wood argued that the convicts were victims of economic hardship, petty offenders or political prisoners.⁹⁷ There were significant numbers of political

93 Péron quoted in n 86 above, 11.

94 Arago quoted in n 86 above, 27.

95 Maçon Society essay competition of 1827.

96 Forster, n 86 above, 115.

97 F. Watson, *The Beginnings of Government in Australia* (Sydney: Library Committee of the Parliament of the Commonwealth, 1913) and G. A. Woods, 'Convicts' (1922) 8 *The Royal Australian Historical Society, Journal and Proceedings* 177–208.

prisoners, at most 4000,⁹⁸ but they were a very small proportion of the 187,000 convicts who came out. There were convicts transported for very minor first offences, even for stealing a prayer book. There were also some murderers. But neither serious violent offenders nor petty first-timers were the mainstream. The majority, men and women, were sent out for middling to minor property offences without violence. Only 38 per cent of those who arrived between 1788 and 1853 were first offenders; only 11 per cent for those who went to Tasmania.⁹⁹

Hancock, Clark, Shaw and Robson rightly reacted against the romantic notions of Watson and Wood,¹⁰⁰ but Manning Clark in particular, overreacted. Clark characterised a large proportion of the convicts as professional criminals, which was clearly not the case given what we now know both about the convicts and the nature of English crime.¹⁰¹ Then came the revision of the revision of the revision with *Convict Workers*, an analysis of convicts transported to New South Wales between 1817 and 1840.¹⁰² Nicholas and Shergold reached a conclusion which seems unexceptionable in respect of most (though hardly all) of the convicts: 'They were not professional and habitual criminals, recruited from a distinct class and trained to crime from the cradle'.¹⁰³ However, without being a 'professional criminal' or a member of a 'criminal class' one can be a serious repeat offender and many of the convicts were. Shlomowitz seems justified in questioning Nicholas and Shergold for offering a different interpretation of the origins of the convicts 'without evaluating the evidence used by Clark, Shaw and Robson in support of their interpretation' and for tending 'to take quantitative evidence at face value and to eschew the use of qualitative evidence'.¹⁰⁴ Even at the quantitative level, it is somewhat shocking that they take their own non-random 1817-40 NSW sample so seriously while Robson's random 1787-1852 sample of cases from all colonies is dismissed.

At a time when crime rates in England were undoubtedly at an historically high level,¹⁰⁵ the transported convicts were essentially that group of offenders whose crimes contemporaries believed to be not serious enough for the scaffold, but too serious for more minor options like the stocks, flogging, fines or a warning following a period of confinement served up to the time of trial. Transportees were fewer than a quarter of those convicted.¹⁰⁶ Often, as today, they had committed a number of crimes and the one for which they were convicted was the one that gave the authorities the sentencing result they wanted. In the late eighteenth century transportation was the standard sentence for remitted capital crimes.

Between 1823 and 1827 Peel began replacing the death penalty with transportation for various property offences, partly because juries were refusing to convict under the shadow of the gallows. The beneficiaries of these liberalising laws went to

98 G. Rudé, *Protest and Punishment: The Story of the Social and Political Protesters. Transported to Australia, 1788-1868* (Oxford: Clarendon Press, 1978).

99 Robson, n 89 above.

100 Hancock, n 89 above; Clark, n 89 above; A.G.L. Shaw, 'The British Criminal and Transportation' (1952) 2 *Tasmanian Historical Research Association, Papers and Proceedings* 29-33, Robson, n 89 above; Watson, n 97 above; Woods, n 97 above.

101 Neal, n 36 above, 50-54.

102 Stephen Nicholas (ed), *Convict Workers: Reinterpreting Australia's Past* (Sydney: Cambridge University Press, 1988).

103 Stephen Nicholas and Peter R. Shergold, 'Unlocking the Past' in S. Nicholas (ed), *Convict Workers: Reinterpreting Australia's Past* (Sydney: Cambridge University Press, 1988) 7.

104 Ralph Shlomowitz, 'Convict Workers: A Review Article' (1990) 30 *Australian Economic History Review* 82, 84; *ibid.*

105 n 82 above, 111.

106 Shlomowitz, n 104 above, 84.

Australia instead. This suggests they would have been mostly middling criminals. Many of them had a level of involvement in crime rather like those who sit in our prisons today. On the one hand, the transportees were less serious offenders because the worst of them were still culled by the scaffold. On the other hand, they were selected from a much higher crime society than today. The majority who were convicted repeat offenders were probably serious recidivists. How could we make this speculation? Well even today self-report studies suggest that the odds of a crime by a young offender leading to a conviction are only about one in a hundred. So the chances that someone who has committed only two offences will be convicted for both of them is about 1 in 10,000. That is today. But in the era before professionalised policing the odds of crime being undetected were much more favourable to the criminal. It follows that most late eighteenth and early nineteenth century offenders convicted of more than one offence would have committed many offences. This does not mean they belonged to a criminal class or were professional criminals. Among them would certainly be some generally honest dead unlucky first or second time offenders. As with criminal statistics today, we can never be sure how serious an individual criminal is from the official record of her or his offending. Contemporary data suggest that official records of crime correlate positively with more inclusive ways of counting crime and we can only assume, without evidence, that this would likely hold true for 200-year old data. Most criminologists, I suspect, would react as I do to reading the data on the scant criminal histories of the convicts. There is not evidence that these were dangerous people, members of a criminal class. Nor can we conclude romantically that they were a combination of poachers victimized by a cruel aristocracy, waifs stealing loaves of bread to feed their brothers and sisters, trade unionists and political prisoners. Rather to the criminologist's eye they look like a typically disparate tranche of convicted criminals – some of them honest and unlucky folk, some vicious criminals who were lucky not to have been caught for scores of more serious crimes, most of them in-between. They constitute an appropriate data set for testing theories of crime prevention. Certainly the exclusion of most murderers (who went to the scaffold) from the data set does not render them an easy group to reform; on the contrary, murderers have lower re-offending rates than the common property offenders who dominate the Australian convict data.

Why Do the Australian Historians Think Emancipist Families Became Mostly Law-Abiding?

Most of the great Australian historians seem to have been persuaded by the fact that the English Tory, Whig and Radical detractors of transportation argued from abstract reasoning, especially Benthamite reasoning.¹⁰⁷ In contrast, the Australians who argued that transportation was working in reforming criminals argued from rich empirical experience. They told concrete stories of reformed and flourishing law-abiding lives. Most English commentators tended to dismiss the Australian testimony as self-serving or debased opinion injured by the profits and degradation that came from dominating other men as slaves.

But this did not ring true to Australian historians who read the biographies of these men. Some of them had retired to England at the time of their testimony and

¹⁰⁷ Their arguments were revealed in documents such as testimony before British parliamentary enquiries.

had no ongoing financial interests at stake. More importantly, the Australian believers in the reformatory accomplishments of transportation were such different men: exclusives and emancipists, judges and governors, businessmen and farmers. Men like King, Arthur, Macquarie, Bligh, Darling, Forbes, Wentworth and Macarthur were so ideologically and philosophically disparate, so often hated one another, yet could agree on their empirical observations of this matter at least. Obversely, the Australian historians tended to dismiss the alarmist testimony of James Mudie about sodomy, sloth, uncontrollable stealing and violence, even that 'all' the convict women were prostitutes, as self-vindicating testimony of a tyrant who had treated his convicts so brutally that they conspired to murder him.¹⁰⁸ Ironically, it was Mudie's stories of depravity which moved those English commentators concerned about the effects of tyranny and slavery.

The centrality of this question of who were the self-serving observers and who were not is perhaps why another kind of evidence that impressed some of the Australian historians was the observations of visitors with independence of judgment who saw the colony first hand. The most frequently cited assessment is that of Charles Darwin who stopped at Sydney Cove on *the Beagle* in 1836: 'As a means of making men outwardly honest, of converting vagabonds, most useless in one country, into active citizens of another, and thus giving birth to a new and splendid country, it has succeeded to a degree unparalleled in history'.¹⁰⁹ It is also worth noting that on the same voyage Darwin was similarly impressed with the accomplishments of Indian convicts banished for life to Mauritius, mostly for gang robbery and murder.¹¹⁰

While it was stories of convicts' lives that influenced most historians, some did cite the informal records of masters of convicts such as Potter Macqueen, who claimed that 'of his servants, sixty-three had been reclaimed and received their tickets, one reclaimed had returned to England, and sixty-two though still under sentence were well-conducted; compared with these, twenty-nine were indifferent, seven depraved, eleven had been sentenced to the gangs or penal settlements, and seven, though freed, were worthless. Apparently 126 were reformed ...'.¹¹¹ These kinds of data are not very persuasive and may indeed have been self-serving.

Hughes cites records of Judge W.W. Burton on the backgrounds of the 827 men he tried between 1833 and 1838 to show that only 4 per cent of them were Currency Lads and none of these had committed murder or grand larceny.¹¹² Hughes uses these data to support the widely reported claim of contemporaries that the descendants of the convicts were a low-crime group. Hughes also looked at all convictions for indictable offences in 1851,¹¹³ concluding that only 6 per cent of them were of Currency Lads and Lasses and that the number of convictions per 100,000 in 1861 was a tenth of the 1831 level.

Ward cites a somewhat different kind of evidence when he quotes from a letter from the Convict Department to the Colonial Adjutant-General assessing the fate of the 60,000 prisoners who had been transported to New South Wales:

108 James Mudie, *The Felony of New South Wales* (London: Whaley and Co, 1837).

109 Quoted in n 35 above, 244-45.

110 Stephen Nicholas and Peter R. Shergold, 'Transportation as Global Migration' in S. Nicholas (ed), *Convict Workers: Reinterpreting Australia's Past* (Sydney: Cambridge University Press, 1988) 33.

111 Shaw, n 35 above, 245.

112 n 65 above, 356-57.

113 *ibid.*, 587-88.

38,000 are now filling respectable positions in life, and earning their livelihood in the most creditable manner ... Of the residue, death and departures from the colony will account for the greater part; and I am enabled to state that only 370 out of the whole are now undergoing punishment of any kind.¹¹⁴

This is so low a rate of re-offending that it is hard for any criminologist to believe.

Perhaps a more persuasive kind of evidence that Australia was a low crime society from the mid-nineteenth to the mid-twentieth century is the persistent reports of citizens feeling no need to secure their property. Ward quotes an observer of the Taron diggings where a high proportion of the diggers 'must still have been ... old hands':

[N]o one scarcely remained to take care of their tents during the day ... [and] ... at night no one ever thought of taking their mining tools away from their claim, and I scarcely ever heard of any being stolen, and no one ever made any hesitation of lending another a crowbar or anything they wanted, no matter whether stranger or not, it was sure of being returned.¹¹⁵

Through the 1950s one could tell the same story of the tent cities littered with fishing tackle on Australian beach-fronts; indeed it was widespread for houses in both the city and the bush to be left unlocked until then.

From an even earlier period, Edward Eyre had a property from 1834 on the Molonglo Plains nearby where I am writing at the Australian National University.

I have constantly known two convicts sent down to Sydney quite by themselves, a distance of 200 miles in a dray full of wool drawn by oxen, and having, after depositing their wool at the merchants', to bring back a load of ... clothing, flour, tea, sugar, tobacco and other groceries – luxuries for the master, and even wines, beer and spirits, and yet tho' this journey involved an absence of five or six weeks during which the men were tempted constantly by the presence of so much property and the facility of appropriating it, the instances were very rare in which any plundering took place or loss ensued.¹¹⁶

Sir Keith Hancock supports his assertion that convict descendants became law abiding by pointing out the lower crime rate in Tasmania compared to other jurisdictions which received fewer convicts.¹¹⁷ Hancock sums up by asserting 'the attested fact that the Australian population does not fall below the high average of respect for law which is attained in British communities'¹¹⁸ This is probably right as a statement of official crime rates in 1930, which were hitting their all time lows in Australia.¹¹⁹

Henry Reynolds more systematically explored the Tasmanian thread of Hancock's argument four decades later.¹²⁰ Manning Clark also picked it up, asserting, without citing any evidence, that crime was on the decrease in Tasmania by Governor Arthur's departure in 1836.¹²¹ Reynolds shows that in the 1857 census, 50 per cent of the adults and 60 per cent of the adult males in Tasmania were still convicts or ex-convicts. The percentage of those convicted of serious crime who were convicts/emancipists fell from 93 per cent in 1848–49 to 70 per cent in 1866–67, to 44 per cent in 1875. The latter numbers are actually not bad in

114 Ward, n 28 above, 30.

115 *ibid* 114; see also Clark, n 89 above, 274.

116 Edward Eyre quoted in Hughes, n 65 above, 314.

117 Hancock, n 89 above, 40–41.

118 *ibid* 40–41.

119 Satyanahu K. Mukherjee, *Crime Trends in Twentieth Century Australia* (Sydney: Allen and Unwin, 1981).

120 Henry Reynolds, '“That Hated Stain”: The Aftermath of Transportation in Tasmania' (1969) 14 *Historical Studies* 19–31.

121 Clark, n 79 above, 310.

terms of the reform of the convicts. When we bear in mind how disproportionately male the convicts were compared to the rest of the population (combined with the much higher rates of offending by males than females in all societies) we can account for much of their overrepresentation by sex alone. On the other hand, one would think the emancipists were getting to be too old for crime by 1875. Yet at that time Tasmania still had an imprisonment rate higher than any nation in the world has today (630 per 100,000).

The more revealing comparison in the Reynolds data is that, as one would expect, between 1857 and 1864 Tasmania had substantially higher numbers of Supreme Court convictions per 1,000 population compared to New South Wales, Victoria and South Australia.¹²² As Wakefield would have predicted, the Tasmanian crime rate was almost three times as high as that of the wholesome citizens of South Australia.¹²³ What then seemed to happen was a much more dramatic fall in the Tasmanian crime rate than occurred in the other colonies. Tasmania had 44 hangings from 1856–65 and 4 from 1866–75. It seemed to completely miss the second wave of bushranging that plagued the mainland colonies in the 1860s and 1870s. By 1889 the Tasmanian imprisonment rate had plummeted to be below average by international standards; by the 1911–1920 decade it dropped further to be 30 per 100,000, lower than *any* developed country in the world today.¹²⁴ The reason, Reynolds shows, is that between 1883 and 1887 supreme court convictions per 1,000 population were 90 per cent *higher* in Eastern Australia as a whole than in Tasmania.¹²⁵ A more impressive accomplishment because Tasmania missed the gold rush and was by far the least attractive destination for free immigrants of all the antipodean colonies. Tasmania had been transformed from being by far the most crime-ridden colony to by far the least crime-ridden. In fact, Tasmania became incredibly peaceful. There were only 22 convictions for homicide in Tasmania in the decade 1875 to 1884. There then follow 32 years without a homicide conviction; the next one is 1916!¹²⁶ There were 25 rape convictions in 1875–84, but 16 years to the next conviction in 1900; 59 robbery convictions in 1875–84, 3 in 1885–94, none in 1895–1904.¹²⁷

There may be many reasons for this. Tasmania may have missed out on the Kelly gangs of the late nineteenth century because the banks most worth robbing were in Victoria. Reynolds does not discuss this possibility.¹²⁸ However, I do think the possibilities he does mention are more interesting and plausible. It seems that Tasmania always recognised that it had a problem with the kind of people providence provided to run its economy and government. It had more convicts and more of the very worst convicts and it kept receiving them for longer than anywhere else. And it had less hope than other colonies of making up for this limitation on its human capital through immigration. So it invested in social welfare and policing of a sort that would requalify its convicts as good citizens. Reynolds shows that Tasmania spent more on prisons and other charitable institutions than the other colonies. Such expenditure accounted for 17 per cent of the Tasmanian budget in 1866, a decade after the end of transportation, compared

122 n 120 above, 21.

123 Wakefield advocated the idea of South Australia as a colony founded by free settlers.

124 Calculated from Satyanahu K. Mukherjee, Anita Scandia, Dianne Dagger and Wendy Matthews, *Source Book of Australian Criminal and Social Statistics* (Canberra: Australian Institute of Criminology, 1989).

125 n 120 above.

126 n 124 above, 440. There were some murder trials but no convictions.

127 *ibid* 445.

128 n 120 above.

with 6 per cent in New South Wales and 4 per cent in Victoria. Tasmania was poorer and had less to spend than the booming mainland colonies, but in 1866 it spent £305 per 1,000 people on charities compared to £138 per 1,000 people in the mainland states. Reynolds concludes that Tasmania 'gave birth to welfare policies which were probably the most advanced and certainly the most expensive in Australia'.¹²⁹ In addition, the police paid regular visits to ticket-of-leave holders and emancipists at their homes. If their intelligence was that they were straying from the straight and narrow, the police let them know they were being watched. We cannot reconstruct what the welfare-surveillance strategies were that worked so well. But we can be sure that there was a prodigious investment in making them work and that their working was not only reflected in a plummeting crime rate. The consumption of spirits also fell to be exactly half the level of Eastern Australia as a whole by 1889. It was 0.69 gallons per head in 1889, having been 2 gallons a head in the Tasmania of 1857.¹³⁰

In the radically different context of neonatal intensive care units in contemporary Chicago, Heimer and Staffen have developed a theory that gives a good account of why this should have happened in Tasmania and indeed of why conditions of labour shortage in Australia as a whole should have led to the successful reintegration of the convicts.¹³¹ Heimer and Staffen attack my claim that the US is an individualistic society that regulates through stigmatisation, particularly with respect to its underclass.¹³² The critique is advanced by revealing a context where American social control is highly reintegrative, especially toward the disadvantaged. This context is the social control of mothers of babies in intensive care units. In this setting, the medical staff need the mothers to do their caring well, to be attentive rather than neglectful with their babies. If they don't teach the mothers to be 'good' mothers, the baby will recurrently come back to the state with renewed intensive care needs. Systematically, the Heimer and Staffen data show that disadvantaged mothers are less likely to suffer stigmatising regulation and more likely to be dealt with in a reintegrative fashion.¹³³ White married mothers were actually more likely to be dealt with stigmatically when they neglected their baby's demanding care needs than were unmarried teenage African American mothers. Heimer and Staffen developed the following more general theory of labelling (stigmatisation):

- (1) labels are not always applied disproportionately to disvalued groups; (2) when excluding deviants is costly, a group, whether a society, community, or organization, will be more reluctant to label anyone as deviant; and (3) when a group cannot function well without the participation of its deviant members, it will work harder to rehabilitate them.¹³⁴

Australia needed the convicts, so instead of labelling them as convicts, "government men" were invited to the governor's table (at least some of them). In circumstances of labour shortage where most of the adults were convicts, Australia could not afford to give up on them.¹³⁵ It had to work hard at reintegrating rather than stigmatising them. All these imperatives were simply much more profound in Tasmania because it had less access to free labour and more difficult convicts to contend with.

129 *Ibid* 30.

130 *Ibid* 22.

131 n 1 above.

132 *Ibid* Braithwaite, n 82 above.

133 *Ibid*.

134 *Ibid* 636-37

135 Indeed this was institutionalised in New South Wales: 'no convict was to be assigned to a master who frequently returned convict servants especially for trifling offences, without making efforts to reform them...' (Clark, n 79 above).

Reintegration worked everywhere; it simply worked better in Tasmania because the Tasmanians had no alternative but to struggle harder at it.

While the fall in the crime rate was not as sharp in New South Wales as in Tasmania, thanks to Sturma we have superior data for New South Wales.¹³⁶ Figures 3 and 4 show the decline from the last decade of transportation (the 1830s) until 1861 in convictions before Superior Courts (which means convictions for the most serious crimes). It was a steep decline for all major offence types. Disaggregating further by offence type, we find that cattle stealing was the property offence that declined most sharply. At its peak in 1835, cattle stealing convictions occur at more than forty times the rate of the low point in 1850.¹³⁷ Note that with cattle stealing, all of the decline occurs before the gold rush, and with all the other comparisons most of the decline occurs before the gold rush. Hence we cannot interpret these falls as a result of the change in the composition of the population occasioned by the surge in free immigration prompted by the gold rush, though admittedly there was also a sharp rise in free immigration in the 1830s and 1840s. While it may not be immediately obvious from scanning Figures 3 and 4, the steepest decline is for offences against property with violence. At the 1834 peak the rate there is more than twenty times the rate at the 1857 low. Assault, which is not recorded separately in Figures 3 and 4, also declined more sharply than property offences.¹³⁸ From its peak in 1839 to its lowest point at

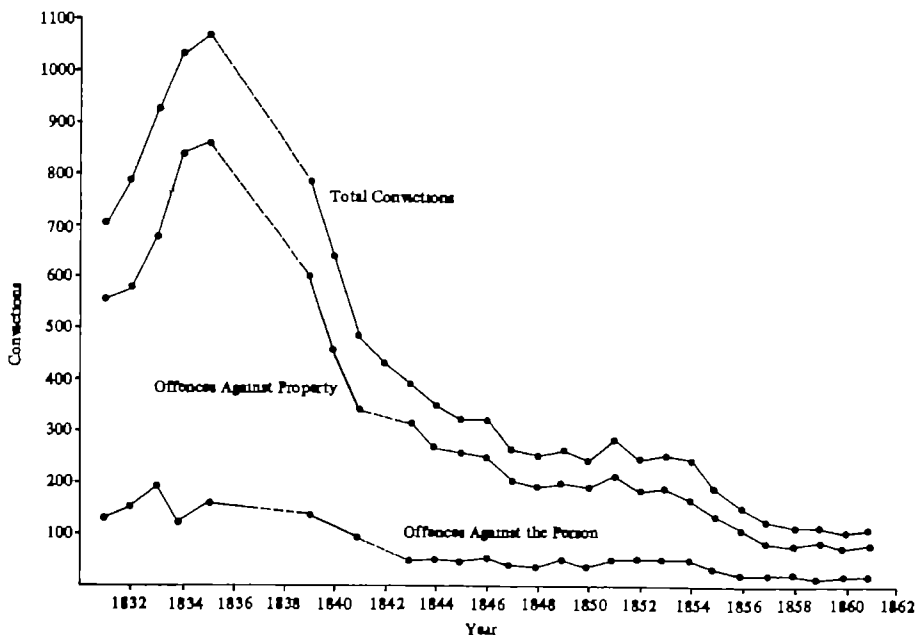


Figure 3: Total Convictions, Convictions for All Offences Against Property and Convictions for All Offences Against the Person per 100,000 Inhabitants, NSW Superior Courts, 1831–61.¹³⁹

¹³⁶ Sturma, n 70 above.

¹³⁷ *ibid* 104.

¹³⁸ *ibid* 94.

¹³⁹ *ibid* 66.

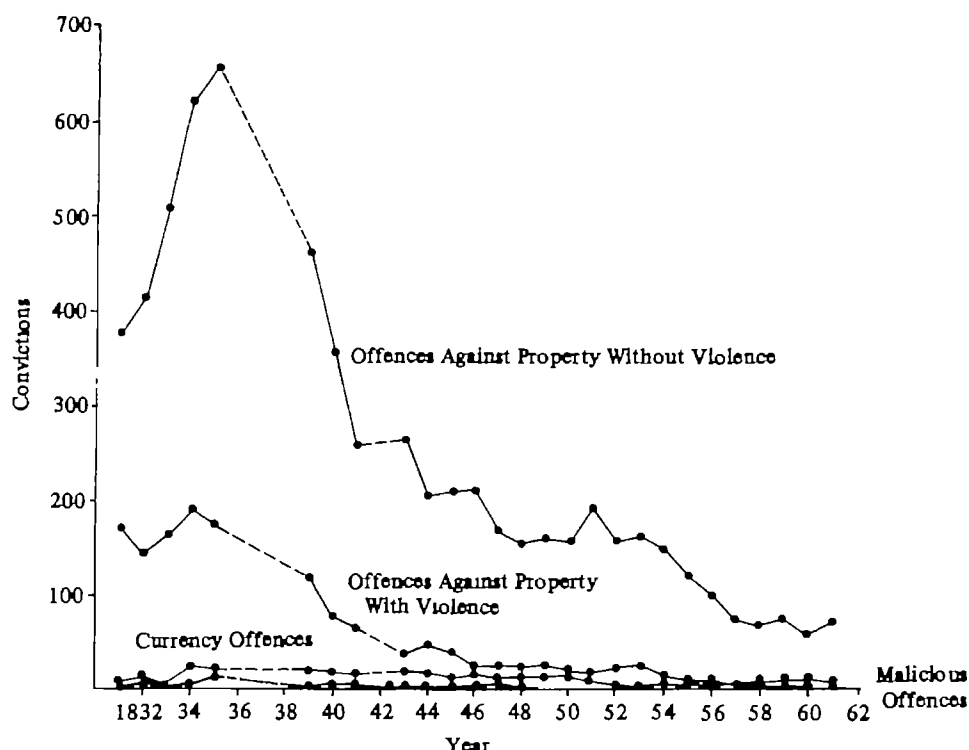


Figure 4: Convictions for Offences Against Property Without Violence, Offences Against Property With Violence, Offences Against the Currency and Malicious Offences Against Property per 100,000 Inhabitants, NSW Superior Courts, 1831–61.¹⁴⁰

the end of the time period, murder and manslaughter declined at a rate similar to that of property offences. We should take special note of the homicide rate because it is the most serious offence and the most reliably measured: definitions of what constitutes stock theft worthy of prosecution may change over time and the capacity of the police to detect it change as the police become more professional. At all places and times in history, however, when a corpse turns up with a knife in it, this fact tends to be noticed and investigated.

Grabosky's data for Sydney for the decades preceding Sturma's data suggest that violent crime declined from Macquarie's assumption of the governorship in 1810 and continued to fall until Sturma's data pick up the trend.¹⁴¹ Property crime per 1,000 inhabitants seemed to rise markedly in the 1820s, however.

Grabosky's data also take us beyond the Sturma series. Grabosky concurs with Sturma's judgement that the most reliable measure, albeit with profound

¹⁴⁰ *ibid* 67.

¹⁴¹ Peter Grabosky, *Sydney in Ferment. Crime, Dissent and Official Reaction* (Canberra: Australian National University Press, 1977) 51–68.

limitations, is Supreme Court convictions. Grabosky finds that both violent and property crime continue to decline after 1860, but much more slowly than the sharp declines Sturma documents between 1830 and 1850. Basically crime rates per capita move inexorably but gradually further downwards until the 1930s after which they move gradually upwards. But it never returns to anything like its peak levels. The odds of being murdered in New South Wales today are about 15 times as good as they were at the peak homicide rate of the 1830s. Indeed you were about three times as likely to be murdered in Sydney then compared to the odds of being murdered in a large US city today. We don't know what the homicide rate was in Sydney before the 1830s, before the clear decline in violence that occurred under Macquarie, but it must have been a genuinely dangerous place.

Reynold's data suggest a steeper drop in the other colonies between 1860 and 1887 than is evident in Grabosky's New South Wales data.¹⁴² The fall in Supreme Court convictions in Tasmania is 80 per cent; for Victoria, South Australia and New South Wales combined it is 50 per cent. During this period, during and following the American Civil War there is a sharp upturn in US crime, probably mainly African-American crime.¹⁴³

Grabosky rightly points out that the decline in the maleness of the colony during the century must have been a major factor in the extraordinary decline that occurred.¹⁴⁴ This would also account for why violent crimes (which are more disproportionately male crimes than property offences) declined more sharply than property crimes. The biggest declines in the maleness of the colony, however, occurred before the Sturma series begins in 1831 and the extent of the fall is not something that sex differences in offending rates could begin to wash away. Sturma points out that in 1841 convicts and ex-convicts made up 35 per cent of the population but 70 per cent of the persons tried.¹⁴⁵ Convicts having only twice the crime rate of the free is a remarkably low ratio when one considers that the proportion of women among the convicts was so low. The male and female convicts do indeed seem to have become progressively more law-abiding and reared children who pioneered a low-crime future.

As with most historical research on crime, it is impossible to say whether various crimes of the powerful got worse or better in the course of the 19th century. We can be fairly sure that murder of and theft from Aboriginal people was worst earlier in the century. By the end of the century most of the land that was taken through violence had long since been alienated. In our crucial case, Tasmania, by 1818 the Aboriginal population had already halved¹⁴⁶ and in 1876 the last "full-blood" died. In the second important case, New South Wales, it had halved by 1840 and halved again between 1840 and 1860.¹⁴⁷ For Australia as a whole, while massacres of Aborigines continued to be a fact into the twentieth century, most of the decimation of the Aboriginal people occurred in the first half of the nineteenth

142 n 120 above, 22.

143 M.S. Hindus, *Prison and Plantation: Crime, Justice and Authority in Massachusetts and South Carolina, 1767-1878* (Chapel Hill: University of North Carolina Press, 1980); Ted R. Gurr, 'Historical Trends in Violent Crime: A Critical Review of the Evidence' in M. Tonry and N. Morris (eds), *Crime and Justice. An Annual Review of Research*, Vol. 3 (Chicago: University of Chicago Press, 1981); E. H. Monkkonen, *Police in Urban America, 1860-1920* (New York: Cambridge University Press, 1981) 76-77.

144 n 141 above, 51-68.

145 n 70 above, 77.

146 Lyndall Ryan, *The Aboriginal Tasmanians* (St. Lucia: University of Queensland Press, 1981) 79.

147 L.R. Smith, *The Aboriginal Population of Australia* (Canberra: Australian National University, 1980) 107.

century, and by the 1880s in all or most states the Aboriginal population was no longer falling.¹⁴⁸ While the peak period of white on Aboriginal violence coincides with the peak period of white on white violence (the first four or five decades after settlement), I do not want to imply that the theory of convict reintegration in this essay is also a good theory of genocide.

In summary, we might say that none of the strands of evidence on which the Australian historians rely in reaching their conclusion is very persuasive by the standards of twentieth century criminology. Putting the strands together, however, the most plausible reading of the evidence is that emancipist families did become rather law-abiding. In the Tasmanian case, this seems to have happened to a stunning degree.

Damned Whores and God's Police

Mary Reibey's picture appears on the back of Australia's \$20 note. She was a convict who became a business leader, owner of five ships, seven farms, warehouses and numerous buildings in the centre of Sydney. Australians take pride that in few countries could a female convict shake her chains to become a major figure in its mercantile class. But Mary Reiby achieved this by getting her start through wedlock. She married a free immigrant with a business. When he died she built that business into a major trading operation. Absent a strategic marriage and Mary Reiby might not have secured a certain place in Australian history. Admittedly, women enjoyed greater equality of access to the law than in England.¹⁴⁹ Yet while wives did have convict husbands assigned to them to run small businesses like bakeries, the practical reality seems to have been that the husband was the master, the wife only legally so. Many male convicts became business leaders; perhaps Mary Reiby was the only female convict who made it to the top in the business world. Women got fewer opportunities through assignment than men. Assignment to work on the land was rarely available to them. While many secured positions as domestic servants, the stereotype of convict women as uncontrollable whores caused masters to prefer male servants.

I have not found how to disaggregate the data in the last section on crime rates by sex. Nor have I seen data on how many female convicts were given land. We can be fairly sure, however, that little land was given to women emancipists. Men were given 50 acres if they had a wife to support, only 30 if they were single. And they got another 10 acres for each child. The policy was clearly intended to encourage marriage as part of the strategy of reform through reintegration. Pursuant to it, women were denied economic security by any other route than marriage. Various Governors also attempted to enforce policies to deny/withdraw land and convicts to men who lived with 'whores', a situation many of the men preferred over marriage.

There is insight in Anne Summers' celebrated analysis that Australian women were first 'damned whores' and then 'God's police'.¹⁵⁰ Interestingly, it is affirmed by the French histories of the colony recently revealed by Colin Forster.¹⁵¹ And whoredom was one respect where the Tory and Whig damnations of transportation were not so wide of the mark. Most of the convict women were not stigmatised as

148 n 147 above.

149 Kercher, n 62 above.

150 Anne Summers, *Damned Whores and God's Police* (Ringwood: Penguin, 1994).

151 n 86 above.

whores in England. Even the one fifth who are officially recorded as prostitutes in Robson's data is problematic (see Oxley).¹⁵² Many of them may well have been so classified because they were servants who slept with their master. Note also that prostitution itself was not a transportable offence.

It is difficult to ascertain how many had been prostitutes before coming to Australia: Robson calculates that about one-fifth had engaged in full- or part-time prostitution. So the wholesale adoption of whoredom on coming to Australia has to be explained in terms of the social climate of the country and the expectations held of women. It was deemed necessary by both the local and the British authorities to have a supply of whores to keep the men, both convict and free, quiescent. The Whore stereotype was devised as a calculated sexist means of social control and then, to absolve those who benefited from it from having to admit their actions, characterized as being the fault of the women who were damned by it.¹⁵³

Caroline Chisholm, pictured on Australia's \$50 note, led a social movement and a programme of practical help to protect newly arriving women – free and convict – from whoredom.

If Her Majesty's Government be really desirous of seeing a well-conducted community spring up in these Colonies, the social wants of the people must be considered. If the paternal Government wish to entitle itself to that honoured appellation, it must look to the materials it may send as a nucleus for the formation of a good and great people. For all the clergy you can dispatch, all the schoolmasters you can appoint, all the churches you can build, and all the books you can export, you will never do much good without what a gentleman in that Colony very appropriately called 'God's police' – wives and little children – good and virtuous women.¹⁵⁴

Chisholm established, with public subscriptions, what amounted to a national job placement service. Her intelligence networks especially looked for employment opportunities in the bush. But she was not looking for permanent jobs in her 'matrimonial excursions in the Australian bush'. 'Whenever she found a comfortable farm owned by a reputable bachelor or widower, she placed a suitable woman with the nearest married neighbour and "in the natural course of events many suitable and happy marriages were the result"'.¹⁵⁵

The systematic research of twentieth century criminology does suggest that Caroline Chisholm was right that a happy marriage and a secure job are the two most important life course changes that end criminal careers.¹⁵⁶ It is likely that the women Caroline Chisholm sent all over Eastern Australia to be God's police did play a significant role in civilizing the wild colonial boys.

The Australian legend which the convict identity helped forge was a patriarchal one. Mateship was about male egalitarianism. It etched a culture that remained robustly patriarchal through the twentieth century. At the same time, the excess of being both damned whores and God's police in a colony tamed through their suffering was a crucible that created women of special stuff – Reiby and Chisholm among them. More importantly, it was a crucible for as inspiring a women's

152 Robson, n 89 above. See Deborah Oxley, 'Female Convicts' in S. Nicholas (ed), *Convict Workers: Reinterpreting Australia's Past* (Sydney: Cambridge University Press, 1988) 85–97.

153 Summers, n 150 above, 332.

154 Caroline Chisholm in 1847 quoted in n 150 above, 337.

155 *ibid* 347.

156 Robert J. Sampson and John H. Laub, *Crime in the Making: Pathways and Turning Points Through Life* (Cambridge, Mass.: Harvard University Press, 1995); Elmar G. M. Weitekamp, Hans-Jürgen Kerner, Wolfgang Stelly and Jürgen Thomas, 'Desistance from Crime: Life History, Turning Points, and Implication for Theory Construction in Criminology' in K. Baasman and S. Karstedt (eds), *Social Dynamics of Crime and Control: New theories for a World in Transition* (Aldershot: Dartmouth, 2000).

movement as any nation can claim. A letter to the *Bulletin* signed 'A New Woman' in 1895 captures the way a feminist identity was forged in the fires of being damned whores and God's police:

What we want is not less parental responsibility, but more, and the great aim of the Woman Movement is to secure equal justice between man and woman, and to uplift the sacred responsibility of parenthood, which has too long been sacrificed to the insatiable Moloch of Lust.¹⁵⁷

A pre-eminent leader of the first wave women's movement was Louisa Lawson, founder of the feminist newspaper *Dawn*. She was the mother of Henry Lawson, the most influential literary interpreter of the egalitarianism of male mateship in the Australian bush. An egalitarian ethos, even if it is a racist and patriarchal one, is still a resource for those who want to open new frontiers of equality. It is hard to build momentum for egalitarian social movements in nations where appeals to equality have no cultural resonance. In contrast, feminists working in societies where male egalitarianism is strong are in a better position to pull on identity crises that prise open cracks in hegemonic structures (such as contradictions between the position of women and a national identity valuing equality of opportunity).¹⁵⁸

It is in these terms that I think we can understand the paradox of an Australia which is one of the most culturally patriarchal of European societies, yet which has had greater structural equality between men and women as a result of the effectiveness of a vibrant women's movement. The structural equality I refer to is the fact that Australia was the second nation in the world (after New Zealand) to secure votes for women,¹⁵⁹ and that as a result of pulling on a crisis in our identity as an egalitarian nation in the twentieth century it was able to institutionalise a degree of equality between the average incomes of men and women second only to Sweden. The latter accomplishment is currently crumbling as a result of the demise of centralised labour market regulation.

Not surprisingly in a patriarchal culture, Australian women did not do so well in getting elected by the people. Again, the creative institutional response of the Australian women's movement to this has been inspiring – Emily's List (women raising funds to support female candidates) and manoeuvring around absence in the legislature by staking out presence in the bureaucracy (the Australian invention of the femocrat).¹⁶⁰

One of the best ways of changing a society that is patriarchal culturally is to institutionalise change that renders it less structurally patriarchal. The cultural change has been slow in Australia, distressingly prone to set-back, yet arriving generation by generation. This is important to reducing crime because societies that are culturally and structurally patriarchal suffer greater crime especially the hidden crimes of family violence and sexual abuse.¹⁶¹ Women as both God's police

¹⁵⁷ Quoted in n 150 above, 393.

¹⁵⁸ John Braithwaite, 'A Sociology of Modelling and the Politics of Empowerment' (1994) 45 *British Journal of Sociology* 445–79.

¹⁵⁹ Audrey Oldfield, *Woman Suffrage in Australia: A Gift or A Struggle?* (Melbourne: Cambridge University Press, 1992).

¹⁶⁰ Anna Yeatman, *Bureaucrats, Technocrats and Femocrats* (Sydney: Allen and Unwin, 1990). The femocrats in turn cleverly insinuated Australian international leadership through the establishment of the Affirmative Action Agency which was beginning to have success through a creatively cooperative strategy for workplace equality for women before the present government moved to pull its limited teeth.

¹⁶¹ John Braithwaite, 'Poverty, Power, White-Collar Crime and the Paradoxes of Criminological Theory' (1991) 24 *Australian and New Zealand Journal of Criminology* 40–58.

and as feminist institutionalisers of equality have played central roles in rendering the colony of thieves a lower crime society than it otherwise would have been.

Why Was Nineteenth Century Australia Characterised by Male Egalitarianism?

The thesis of Turner is that frontier societies find their identity from struggles on the frontier.¹⁶² Climate and vegetation made the American frontier one of the rugged individualist farmer. In contrast, Ward's Australian frontier was one of holdings that had to be vast to be economically viable and that employed a substantial rural working class to shear the sheep, drove and tend the fences.¹⁶³ The trade union movement and the Labor Party rose to prominence in the bush at first, not in urban factories. The shearer's union, the Australian Workers' Union, was the dominant force. Ward shows that the further outback settlement went, the more the population was dominated by former convicts, men and the Irish.¹⁶⁴ Free settlers, women and the English preferred the security of civilisation to the higher pay in an outback plagued by labour shortages.

The harsh conditions of outback work benefitted from the mateship that aided survival during the convict voyage and assignment. Both were republics of sharing and solidarity in opposition to the powerful (the military in Sydney, the squatters in the bush). From the outset, the tales of this solidarity are legendary. When Governor Hunter compelled the convicts to attend religious services in 1798 the church was burnt down by a convict. Hunter was so furious he offered any informer, even one serving a life sentence, a free pardon and passage home plus a reward of £50. No one turned.¹⁶⁵ Foolish, heroic mates whose legend would be manifest in Gallipoli.

When free male immigrants arrived and confronted the choice between the exploitation of the exclusives or the mateship of the emancipists, there were some good reasons for choosing the latter: 'He will find that he can hardly avoid attaching himself to one party: if to the free, the other will say, Let him alone awhile; the swells will pluck him, and then he will come to us'.¹⁶⁶

There is documentation from as early as 1822 of convicts organizing collectively to withhold labour to enforce demands for higher pay.¹⁶⁷ Convicts were organising as founding members of the Printers' Union while they were still under sentence.¹⁶⁸ Ward argues from a rich fabric of evidence that the mateship of convicts and emancipists was the legend of the outback and that rural working class struggle forged the trade union movement.¹⁶⁹ That trade union movement took male-dominated mateship into the cities. So did the *fin de siècle* literary interpreters of mateship, Paterson and Lawson, and their republican vehicles, *The Bulletin*, the *Queensland Worker*. Paterson's *The Man From Snowy River and other Verses* sold over a hundred thousand copies in the little nation. Lawson came

162 Turner, n 28 above.

163 Ward, n 28 above.

164 *ibid* 66–67.

165 *ibid* 85.

166 James Dixon (1822) quoted in Ward, n 28 above, 36.

167 Ward, n 28 above, 72.

168 Stephen Nicholas and Peter R. Shergold, 'A Labour Aristocracy in Chains' in S. Nicholas (ed), *Convict Workers: Reinterpreting Australia's Past* (Sydney: Cambridge University Press, 1988) 99.

169 Ward, n 28 above.

to be even more widely read; a hundred thousand lined the streets for his funeral. Later C.J. Dennis moved the sardonic genre to the city with two larrikins who the nation took to their hearts, the bloke of *The Songs of a Sentimental Bloke* and Ginger Mick.

A thesis about bush workers in the pastoral industry seems a strange one to advance in a nation that came to be among the world's most urbanised. Its significance can be and has been exaggerated. Yet it is hard to reconcile an utter rejection of it with *Waltzing Matilda* (the sheep thief hounded to his death by police and squatters) and Ned Kelly. It is hard to say there is no convict stain when we see those old diggers on Anzac Day 'illegally' playing two-up, ritually furtive. Gambling of such simplicity it could be played in 'schools' in the prison yard at Newgate by men bound for Botany Bay. In the nineteenth century so many respectable Australians worked to remove the convict stain. Today many progressives are ashamed of the way the solidarity of mates was mobilised to exclude women, vilify Chinese workers who threatened their wage rates and exterminate Aborigines who grafted to share their own land. Denial does not seem preferable to acknowledging all this as part of the complex fabric of a past on which we must build as we heal.

Beyond Foucault

An implication of our Australian story is that the sociology of punishment is excessively obsessed by the emphasis in Foucault's *Discipline and Punish* (and in mainstream penology) on the birth of the prison as the crucial development.¹⁷⁰ Rather than conceive the crucial divide as between prison and punishment inscribed on the body, I want to conceive the opposition between inclusionary and exclusionary regulation as the much more historically fundamental theme.¹⁷¹ It is a deliciously paradoxical one. Stan Cohen's contribution to it was to show how inclusionary theories degenerate into oppressive and exclusionary practices (eg netwidening): 'beautiful theories becoming ugly practices'.¹⁷² The story here is more one of ugly theories (purging England of its depraved – Blackstone's extirpation) giving birth to some beautiful practices. Georgia illustrates the more standard Cohen story of beautiful theories. Here the philanthropist prison reformer James Oglethorpe founded Georgia in 1733 to create a colony that avoided extremes of wealth and that attracted victims of both the prison system and religious persecution. Slavery would be prohibited and larger landholdings forbidden by law. He wanted to help convicts, particularly those imprisoned for debt, to work their way to solvency on American land. Oglethorpe's regulations were defied as greedy planters smuggled in slaves to replace convicts.¹⁷³ The colony which, unlike Australia, set out to be the egalitarian inclusionary republic, was gone with the winds of exclusion.

The inclusion/exclusion divide is so interesting and central precisely because its history is one of inverting itself with profound consequences. So our account of transportation is of an exclusionary move which is transformed into inclusion. Foucault did not take transportation seriously, which is odd because he takes

170 n 2 above

171 Stanley Cohen, *Visions of Social Control* (London: Polity Press, 1985).

172 *ibid.*

173 Paul Johnson, *A History of the American People* (London: Phoenix, 1997) 91.

Bentham so seriously and Bentham took transportation so seriously that he devoted more energy to fighting it to defend his alternative of the panopticon than to any policy idea. In 1802 he published *Panopticon Versus New South Wales*. Nothing could be more offensive to the principles of the panopticon than convict shepherds separated by large spaces from the nearest shepherds; constant surveillance by a flock of sheep was not what Bentham had in mind. Foucault is remarkably ill-informed in the way he dismisses the historical significance of transportation. He says England abandoned transportation 'at the beginning of the nineteenth century'. Transportation to Australia was not abandoned until 1868 and continued to Gibraltar (where 9,000 were shipped) until 1875. There was more transportation to Australia in the second third of the nineteenth century than in the first. More disturbing is the way Foucault fails to document the significance of his own nation's experience of transportation, of the way Tocqueville lost the French debate and French transportation continued to 1938, albeit at lower volume and with less colonial success than English transportation. As late as 1911 transported convicts represented 13.2 per cent of the population of Guiana,¹⁷⁴ though this was a much smaller destination than New Caledonia.

Nicholas and Shergold point out that after 1820 two and a quarter million convicts were transported to destinations that included Australia, Siberia, Singapore, New Caledonia, French Guiana, Gibraltar, the Nicobar Islands, Brazil, Sumatra, the Andaman Islands, Bermuda, Penang, Malacca and Mauritius.¹⁷⁵ Australia is only significant because of its scale (being second only to Siberia as a destination) and because its impacts are the best documented. In the late nineteenth century, Singapore came to be viewed as a success story of building a strong economic infrastructure with convict labour in the way Australia had been so viewed early in the century. Just as the French sent a mission to Sydney in 1802, the Japanese, Siamese and Netherlands governments sent fact-finding missions to the Straits Settlements in the latter part of the century. Throughout the century, Britain threw its weight around to prevent other states following the path to colonial development it had pursued. For example, it resisted attempts by Austria, Italy and Germany to establish penal colonies in the Pacific. France was too powerful for Britain to resist, though Britain did manage to persuade it that a French penal colony in Western Australia would be ill advised. The state of Hamburg actually signed a contract to ship convicts to the Australian Agricultural Company, but the British Secretary of State, Lord Glenelg put a stop to it with the convicts waiting on the ship. Prussia managed to broker a deal with Russia to send some of its convicts to Siberia and Mecklenburg-Schwerin got past the British by sending some to Brazil.

While this transportation after 1820 accounts for the highest volumes, there was also considerable English transportation of convicts to North America in both the eighteenth and seventeenth centuries and some to Africa.¹⁷⁶ There was also some Swedish transportation to New Sweden (Delaware) during the seventeenth century and a momentary Dutch flirtation with transportation to Surinam.¹⁷⁷ During these centuries and throughout the sixteenth and some of the fifteenth century as well, Spain, France, Austria and most Italian and other Mediterranean states banished

174 n 110 above, 35.

175 *ibid*

176 n 35 above, 32–34

177 Pieter Spierenburg, 'The Body and the State: Early Modern Europe' in Norval Morris and David J. Rothman (eds), *The Oxford History of the Prison: The Practice of Punishment in Western Society* (New York: Oxford University Press, 1995) 68.

prisoners to galley slavery, but again in numbers that were modest compared to nineteenth century transportation. Garland points out that both Portugal and Spain used convict labour in their foreign colonies and military settlements from as early as the fifteenth century.¹⁷⁸ In the great European wars of the second half of the eighteenth century, the protagonists used convicts to fight and even took criminals from other countries that did not want them.¹⁷⁹ No fewer than 97 percent of the non-capital convictions decided by Amsterdam's court were to banishment between 1650 and 1750.¹⁸⁰ Similarly in American cities of the same period, the absence of a police force meant that social control operated largely informally so long as undesirables could be excluded: newcomers would be 'warned out' unless they could vouch for their respectability with say a letter from their minister or property.¹⁸¹

A problem with the sociology of punishment has been a tendency to take Foucault too seriously as an historian, especially on this question of the dismissal of transportation. Spierenburg and Finnane can be exempted from this charge,¹⁸² and Garland also at least in passing concedes some importance to transportation in the mid-nineteenth century.¹⁸³ So I submit we have a sociology of punishment too obsessed with the politics of imprisonment to the neglect of the more abstract politics of exclusion and inclusion (of which the politics of imprisonment is a crucial part). An attraction of explanatory theories of exclusion and inclusion is that they map onto normative theories of restorative and retributive justice, a central mapping for those of us who aspire to integrate explanatory and normative theory.¹⁸⁴ From this normative perspective the contest between whether to imprison or to put someone in the stocks is not as central as the contest between whether to seek restoration or retribution. One reason the latter is more normatively central is that it implies asking the question of whether one should punish at all; or rather, restorative justice means stakeholders deliberate in a way where punishment might never be considered as a means of securing symbolic and material restoration. What follows is that regulation is a more fruitful topic of study, in terms of both normative and explanatory theory, than punishment. In this respect, the late Foucault of governmentality has the greater appeal than the Foucault of *Discipline and Punish*.¹⁸⁵

What also follows is that my five-stage history of institutions of regulation is flawed. Between the weak state stage where corporal and capital punishment dominate (the second stage) and the third strong state stage where professional police and penitentiaries dominate we need to interpose a briefer stage where banishment is revived to being a dominant form of punishment, at least for a number of major nations such as England and Ireland, France, Holland, Spain, the Northern Italian States, and Russia. The shift from banishment being a minor form

178 n 9 above, 99.

179 n 3 above, 29–30.

180 n 177 above, 62–63.

181 David J. Rothman, 'Perfecting the Prison: United States, 1789–1865' in Norval Morris and David J. Rothman (eds), *The Oxford History of the Prison: The Practice of Punishment in Western Society* (New York: Oxford University Press, 1995) 112.

182 Spierenburg, n 8 above; Michael Finnane, *Punishment in Australian Society* (Melbourne: Oxford University Press, 1997).

183 n 12 above, 7.

184 John Braithwaite and Christine Parker, 'Restorative Justice is Republican Justice' in Lode Walgrave and Gordon Bazemore (eds), *Restoring Juvenile Justice: An Exploration of the Restorative Justice Paradigm for Reforming Juvenile Justice* (Monsey, NY: Criminal Justice Press, 1998).

185 Garland, n 15 above; n 2 above.

of punishment to being a dominant form occurs at different periods between 1500 and 1900 in different states, depending on where labour shortages occur in the launching of imperial expansion. In short an important partial explanation of the rise and fall of banishment from the home state to the periphery of its imperium is provided by Rusche and Kirchheimer's materialist theory.¹⁸⁶ This means that the resurgence of banishment actually completely overlaps the chronology Foucault favours for his transition from inscribing the body to disciplining the soul.

Just as prisons existed throughout recorded history, so did banishment and so did the infliction of pain on bodies. Figure 2 shows that in nineteenth century Australia, penitentiaries, banishment, corporal/capital punishment and restorative justice all co-existed. My point is one about the chronology of different techniques and institutions of regulation becoming dominant. The conclusion is that the domination of imprisonment is preceded by a brief domination of banishment. Hence the revised sequencing of institutional dominations is restorative justice, corporal/capital punishment, banishment, centralised Benthamite state penitentiaries, Keynesian welfare state probation-prison-parole, and now the contemporary uncertain unfolding of a new regulatory state of neo-liberal governmentalities (community policing, restorative justice, actuarialism, government at a distance).

The era of the domination of the Western debate by transportation was not so long. Yet how can we give centrality to the thinking of Bentham and Tocqueville in that debate without understanding how and why they lost it in their own lifetimes?

Australian transportation was the work of a hegemonic power. Transcontinental shifts of convict labour, especially by Britain and Russia, were of a piece with millions of indentured Melanesians, Chinese and Indians and millions of African slaves who were moved to spaces where labour was scarce as part of empire-building strategies. In the process of penalty being an instrument of imperial expansion, it was transformed in paradoxical ways. It showed us how well restorative justice might work with slum dwellers from the largest metropole. Macquarie's Sydney reinvented what we now describe as restorative justice and Macquarie even used the language of restoration. The colony invented the ticket-of-leave, which was modelled in England and became parole.¹⁸⁷ It established an institution for juvenile offenders and stopped executions in public for the same reasons that these things occurred in England decades later.¹⁸⁸ In its English-driven reaction against the restorative justice of the ticket-of-leave, Australian convict administration produced Edmund Du Cane (1885), who became the driving administrative force and theoretician of severe centralised Benthamite state penitentiaries in Britain. Du Cane was both the author and pre-eminent implementer of the principle of lesser eligibility. At the Norfolk Island penal settlement, Alexander Maconachie had the opportunity to implement the marks system of his new synthesis of progressive movement from confinement to

186 n 3 above, and John Braithwaite, 'The Political Economy of Punishment', in E. L. Wheelwright and K. Buckley (eds), *Essays in the Political Economy of Australian Capitalism*, Volume IV (Sydney: ANZ Books, 1980).

187 Finnane, n 182 above, 162.

188 Governor Bourke ordered that executions be carried out in private after a convict named Jenkins made a famous traditional speech from the drop: 'Well, good bye my lads, I have not time to say much to you; I acknowledge I shot the Doctor, but it was not for gain, it was for the sake of my fellow prisoners because he was a tyrant, and I have one thing to recommend you as a friend, if any of you take to the bush, shoot every tyrant you come across, and there are several now in the yard who ought to be served so' (Ward, n 28 above, 139).

reintegration into the community. This approach was the dominant influence on the US reforms in the aftermath of the 1870 National Congress of Penitentiary and Reformatory Discipline¹⁸⁹ and the Irish system of Sir Walter Crofton¹⁹⁰ and central to understanding how Keynesian welfare state probation-prison-parole evolved everywhere.¹⁹¹ Assignment was copied by the French in their penal colonies as 'assignation'. In the Straits Settlements Indian convicts were given 'tickets-of-leave' on the Australian model and like the Australian emancipists often showed great entrepreneurial flare in running businesses.¹⁹² Doubtless there may be second-order effects here in other parts of the world which are yet to be researched. What we can say is that when Foucault dismissed transportation as something the English abandoned 'at the beginning of the nineteenth century', he dismissed something that mattered.¹⁹³

Conclusion

By the end of the nineteenth century Australia was a low crime society and perhaps as prosperous and egalitarian as any society in the world, certainly in terms of workers' wages, which were higher than in the US and Britain. Where the convict presence was greatest, the fall in the crime rate was steepest. I have argued that mutually reinforcing policies of reintegration and procedural justice toward the convict majority during the first half century of European settlement was one reason for this. This contrasts with another frontier society, the United States, where stigmatisation and procedural injustice toward slaves produced a high crime society. The descendants of the slaves continued to be poor and stigmatised at the end of the century, surrounded by burgeoning wealth beyond their reach. While crime rates of convict descendants seem to have declined to approximate that of the general population, the descendants of the American slaves committed homicide at about eight times the rate of descendants of the free.¹⁹⁴

Australia also ended the century as a patriarchal and racist society with strong traditions of corruption and abuse of power. It had decimated the Aboriginal owners of the continent, the earth, the whales. Moreover, in the twentieth century Australia lost the lead it had held over other countries in prosperity and income equality. As Russel Ward has argued, while the US was a frontier of rugged individualism, the Australian frontier was a collectivist one of working class solidarity informed by convict irreverence for authority. However, it formed a working class that was born to knock those who set themselves above their mates. It was Tories like Menzies, Bruce, Deakin, Fraser and Reid who were seen as born to rule. These men were born to rule by British aristocratic legends, or a bunyip version of them. The monarchy had great resonance with an Australian bourgeoisie who suppressed the convict stain and with women who feared male trade union

189 Some of the Declaration of Principles of the American Prison Association was taken word for word from Maconochie's writing.

190 Eduardo Rotman, 'The Failure of Reform: United States, 1865-1965' in Norval Morris and David J. Rothman (eds), *The Oxford History of the Prison: The Practice of Punishment in Western Society* (New York: Oxford University Press, 1995) 173.

191 John Vincent Barry, *Alexander Maconochie of Norfolk Island: A Study of a Pioneer in Penal Reform* (Melbourne: Oxford University Press, 1958).

192 n 110 above, 33.

193 n 2 above.

194 My thanks to Professor Robert Silverman for accessing US Uniform Crime Report which shows a rate of homicide for African-Americans of 39.1 per 100,000 and a rate of 4.9 for European-Americans.

domination. It was a literate working class which read and talked about politics a lot, but mocked intellectuals. It failed to deliver an egalitarian twentieth century Australia, cut down its own tall poppies, preferring an informal mateship that mocked engaged and disciplined political solidarity. It was more comfortable with defeat at Gallipoli than victory with Fisher, Scullin, Chifley or Whitlam.¹⁹⁵ Then nor would it have tolerated a Hitler, Stalin or Péron. No doubt the light on Australia's hill still has more than a touch of 'convict republicanism'.¹⁹⁶

This work began by positing five stages in the history of regulation: a pre-state stage when restorative justice and banishment were dominant, a weak state stage where corporal and capital punishment dominated, a strong state stage where professional police and penitentiaries dominated, a Keynesian welfare state stage where new therapeutic professions such as social work colonised what became probation-prison-parole, and a contemporarily evolving new regulatory state phase of community and corporate policing (with a revived restorative justice). I conclude that the rise of the penitentiary is preceded by a period when transcontinental movement of convicts was dominant because of its use (like slavery and indentured labour) as an instrument of imperialism. Restorative justice, corporal/capital punishment, banishment, imprisonment and government at a distance have all, in that order, had their period of dominance. Equally they have all been part of the regulatory story of all periods of recorded history.

For the Foucault of *Discipline and Punish*, the penal divide is between disciplining the body versus the soul, scaffold or penitentiary. Australian history illustrates the richer insight from seeing the most fundamentally recurrent historical tension as between inclusion and exclusion,¹⁹⁷ between the normative ideals of restorative and retributive justice. The way we have construed Heimer and Staffen's theory as relevant to both nineteenth century Tasmanian convicts and twentieth century African-American mothers illustrates the value in generality of explanation from researching the inclusion-exclusion opposition.¹⁹⁸ Heimer and Staffen's general claim, which we have supported, is that when those with the power to stigmatise are dependent on the deviant, they opt for reintegration more than stigmatisation.

195 C.J. Dennis's Ginger Mick believed in livin' and lovin' without holding any hope for the future of Australia. But through Gallipoli he reveals larrikinism in the words of Henry Lawson and Manning Clark as 'chivalry - upside down':

An' each man is the clean, straight man 'is Maker meant 'im for,
An' each man knows 'is brother man at last.
Shy strangers, till a bugle blast preached 'oly brother'ood;
But mateship they 'ave found at last; An' they 'ave found it good.

196 J.B. Hirst, "The Australian Experience: The Convict Colony" in Norval Morris and David J. Rothman (eds), *The Oxford History of the Prison: The Practice of Punishment in Western Society* (New York: Oxford University Press, 1995).

197 See Jock Young, *The Exclusive Society* (London: Sage, 1999).

198 n 1 above.

The Wasted Costs Jurisdiction

*Hugh Evans**

The wasted costs jurisdiction is flawed for six reasons, based on an analysis of all reported cases in the last nine years and five years of statistics provided by the Bar Mutual Insurance Fund Limited, and despite the guidance laid down by the Court of Appeal in Ridehalgh v Horsefield [1994] Ch 205. First, it is very costly proportionate to the amount recovered. Secondly, judges can initiate a wasted costs enquiry, which is unfair and even more disproportionately costly. Thirdly, it is procedurally complex. Fourthly, it is unpredictable whether the client will waive privilege, and what the consequences will be whether or not privilege is waived. Fifthly, it is not possible for solicitors and barristers to make contribution claims against each other. Sixthly, it is mostly used against lawyers representing legally aided litigants from whom costs cannot be recovered.

The summary jurisdiction to order a solicitor to pay wasted costs is an old one.¹ However, the original requirement to show serious misconduct meant that it was infrequently used.² The courts decided, after some hesitation, that the (then) new RSC Order 62 rule 11, introduced in 1986, required the applicant to prove only that the respondent solicitor was negligent.³ Since the introduction of sections 4 and 11 of the Courts and Legal Services Act 1990 on 1st October 1991, the jurisdiction has been expressly based, *inter alia*,⁴ on a requirement to show only negligence, and it has been extended to include barristers and other legal representatives as well as solicitors. As a result of these developments, there have been many judgments by the courts on applications for wasted costs, both reported and unreported. Despite the best endeavours of the Court of Appeal, in the leading case of *Ridehalgh v Horsefield*,⁵ to place limits and safeguards on such applications, the jurisdiction is frequently used.

This article will argue that both in practice and in principle, the wasted costs jurisdiction is seriously flawed. Some of the reasons are theoretical, and the practical importance of them may be difficult to judge, and some are practical. The objections are based on an analysis of all the reported cases since the new jurisdiction came into force nine years ago,⁶ particularly those cases heard after *Ridehalgh*. Those decisions suggest, in particular, that the costs of the wasted costs enquiry very often far outweigh the wasted costs which are either sought or

* Barrister, 4 New Square, Lincoln's Inn.

1 For an analysis of the law, rather than critical comment on some of its aspects, see my summary in Jackson and Powell *Professional Negligence* (London: Sweet & Maxwell, 4th ed, 1996) paragraphs 4-102 *et seq* and 5-20 *et seq*, and Flenley and Leech *Solicitors' Negligence* (London: Butterworths, 1999) chapter 11.

2 As far as the inherent jurisdiction is concerned, see *Myers v Elman* [1940] AC 282.

3 See *Gupta v Comer* [1991] 1 All ER 289.

4 There are alternative grounds of improper or unreasonable conduct.

5 [1994] Ch 205.

6 Some 68 cases in all. Section 4 of the Courts and Legal Services Act 1990, which substituted the relevant new provisions relating to costs in s 51 of the Supreme Court Act 1981, came into force on 1 October 1991.

recovered. This impression is strongly reinforced by the figures for the 1994–98 policy years provided by the Bar Mutual Indemnity Fund Limited (BMIF).⁷

Before considering the problems with wasted costs, which I will divide into six areas, it is worth mentioning the competing principles behind the jurisdiction which are clearly set out in *Ridehalgh*:⁸

The argument we have heard discloses a tension between two important public interests. One is that lawyers should not be deterred from pursuing their clients' interests by fear of incurring a personal liability to their clients' opponents; that they should not be penalised by orders to pay costs without a fair opportunity to defend themselves; that wasted costs orders should not become a back-door means of recovering costs not otherwise recoverable against a legally-aided or impoverished litigant; and that the remedy should not grow unchecked to become more damaging than the disease. The other public interest, recently and clearly affirmed by Act of Parliament, is that litigants should not be financially prejudiced by the unjustifiable conduct of litigation by their or their opponents' lawyers.

Even with the safeguards sensibly imposed by *Ridehalgh*, experience and argument suggests that the first interest is still insufficiently met, and that, with respect to the second interest, the remedy is more damaging than the disease.

A costly jurisdiction

It is all too common for the time and costs of the wasted costs enquiry to be wholly disproportionate to the quantum of the wasted costs. It is perhaps redundant to remark that in these Woolfian days, proportionality is a key ingredient in the legal procedure, and it remains to be seen whether the impact of the Woolf reforms will curb the number of wasted costs applications. In *Ridehalgh*⁹ the Court of Appeal commented that:

Material has been placed before the court which shows that the number and value of wasted costs orders applied for, and the costs of litigating them, have risen sharply. We were told of one case in which the original hearing had lasted five days; the wasted costs application (when we were told of it) lasted seven days; it was estimated to be about half-way through; at that stage one side had incurred costs of over £40,000. It almost appears that a new branch of legal activity is emerging, calling to mind Dickens's searing observation in *Bleak House*: "The one great principle of English Law is, to make business for itself . . . Viewed by this light it becomes a coherent scheme, and not the monstrous maze the laity are apt to think it."

The Court of Appeal in *Ridehalgh* was almost certainly referring to *C v C (Wasted Costs Order)*,¹⁰ where the wasted costs hearing eventually lasted fourteen days and cost the parties over £150,000, more than the original costs and far more than the costs which the Court determined had been wasted. This is not new. In *Myers v Elman*¹¹ itself, the leading authority on the inherent jurisdiction to order a solicitor to pay costs personally, the original case lasted five days, and the wasted costs hearing another five days, followed, of course, by appeals to the Court of Appeal and House of Lords.

7 I am very grateful to Redvers Cunningham of BMIF for supplying me with the data. I have been unable to obtain figures from the Solicitors Indemnity Fund. Judging by the reported decisions, it would be likely that there are several times the number of wasted costs applications and orders against solicitors than there are against barristers.

8 n 5 above, 226.

9 n 5 above, 225–226.

10 [1994] 2 FLR 34.

11 [1940] AC 282.

To achieve proportionality, the Court of Appeal in *Ridehalgh* considered that hearings should be measured in hours and not days.¹² Similar sentiments are found in subsequent observations of the Court of Appeal: in *Turner Page Music v Torres Design Associates Ltd*¹³ that the wasted costs procedure is inappropriate if it would result in complex proceedings; in *re Freudiana Holdings Limited*¹⁴ that allegations involving full scale relitigation of the issues in the original trial and which involved dishonesty should not proceed; and in *Wall v Lefever*¹⁵ that the wasted costs jurisdiction should not be used to generate substantial additional costs in satellite litigation which was as expensive and complex as the original litigation. These observations continue to be made because the problem has not gone away.

The length of wasted costs hearings still appears to be disproportionate. Anecdotal evidence from practitioners with considerable experience of wasted costs applications suggests that this is very common indeed. When it comes to reported cases, in some instances one cannot discern the lengths of the hearings, and in many cases both the original and the wasted costs hearings lasted all or part of a day or two, and there is not sufficient information to come to any sensible conclusion about their comparative lengths and costs. The relatively few reported cases which I have found in which the respective length of the hearings can be sensibly deduced and compared appear to support the impression gained from practitioners, which is that the burden of the wasted costs application is all too often disproportionate to what was at stake. Indeed, I have found no reported case where it is clear that the costs incurred in the wasted costs application were justified by the amount of wasted costs sought or recovered.

In *Re a Company (No 006798 of 1995)*,¹⁶ a petition to wind up a company was struck out within a day, but the subsequent (successful) application for the solicitor who swore the affidavit in support of the petition to pay the wasted costs appears to have taken all or part of four days (presumably the last day being judgment and therefore brief), with leading counsel appearing for the respondent. In *Flimlab Systems International Ltd v Pennington*¹⁷ a 'wholly misconceived' application for discovery took one and a half hours. The subsequent wasted costs application was successfully resisted on behalf of the respondent by leading counsel. While the report only indicates that the hearing took all or part of two days, the length and detail of the judgment suggests that it must have taken many times longer than the discovery application. In *Re Freudiana Holdings Limited*¹⁸ the order that the lawyers should show cause was not allowed to proceed at a stage when the respondents' costs had reached £400,000 before the wasted costs hearing itself had been reached. However, the original case did last 165 days, with the respondent lawyers being dismissed on day 85. In *Re Merc Property Ltd*¹⁹ Ferris J. commented that far longer had been spent on the wasted costs application than on the substantive proceedings, and in future applicants for a wasted costs order had to bear in mind the principle of proportionality. In *Warren v Warren*²⁰ the costs

12 n 5 above, 238, a passage repeated and emphasised by Lord Woolf MR in *Warren v Warren* [1997] QB 488, 495.

13 *The Times* August 8 1998.

14 *The Times* December 4 1995.

15 [1998] 1 PCR 605.

16 [1996] 1 WLR 491.

17 [1995] 1 WLR 673. The case was decided before *Ridehalgh* was reported.

18 *The Times* December 4 1995. The figure of £400,000 does not appear in this report, but is found in Flenley and Leech, n 1 above, paragraph 11.8. For a similar recent case see *Chief Constable of Yorkshire v Ainsley* [2000] Lloyd's Rep PN 675.

19 *The Times* May 19 1999. The short report gives no details of the time spent.

20 [1997] QB 488.

wasted concerned a committal hearing which, after a number of adjournments, was dismissed by consent. The wasted costs hearing, which was successful at first instance (but not on appeal), took three days before the Judge, and caused the Master of the Rolls to comment that both the committal proceedings and the wasted costs proceedings were 'totally disproportionate' to the nature of the proceedings involved.²¹

In *Re G and ors (children)*²² the wasted costs hearing took half a day. It was preceded by written submissions, followed by further 'substantial' written submissions and a statement from a solicitor exhibiting a bundle of correspondence, with judgment a few weeks later. While this surprised the judge, one may think that the time spent was modest compared with other cases. However, the application was against four barristers, one a silk, and concerned delays at the hearing caused by two experts not having seen all the relevant documents. In the case of one of the experts, although the judge held that Counsel had acted unreasonably and negligently, he was not confident that any costs had been thrown away, as the wasted time was so insignificant. In the other case, a wasted costs order was made against only one of the barristers, and the costs wasted were those of securing the attendance of one expert witness for half a day. The time spent in the wasted costs hearing²³ was therefore wholly disproportionate to the wasted costs in question. If the point of the wasted costs enquiry, which was initiated by the Court, was to have been to remind the Family Bar in general about their obligations to the Court, that could have been done in a far more economical way.

There are other cases where the wasted costs sought or awarded were specified and were trivial. In *Re Harry Jagdev & Co*,²⁴ the court ordered the respondent solicitors to pay £1,816.09, but the Court of Appeal considered that the order was fatally flawed and quashed it. A similar result was reached in *R v Lawrence*,²⁵ where the Court of Appeal quashed an order for wasted costs in the sum of £1,500. In two cases where the lawyers failed to turn up to a hearing, wasted costs orders were made, £250 in *R v Rodney*,²⁶ and £750 in *R v Secretary of State of the Home Department, ex p Mahmood*.²⁷ The wasted costs in *re Sternberg Rood Taylor & Gill (a Firm)*²⁸ were £1,237.48, and in *Re a Solicitor (Wasted Costs Order)*²⁹ £727.88; both orders were unsuccessfully appealed. If such sums were claimed in ordinary litigation, they would be small claims cases, the costs of fighting them would not be recovered, and it is highly unlikely that anyone would actually think it cost-effective to bring them. Nor would any appeal be made, whatever the result.

The impression derived from the reported cases is that the costs of the wasted costs enquiry often outweigh the amount awarded, although it is based on relatively few cases. The experience of BMIF, which insures all barristers, is much better evidence. Their figures for the five policy years 1994 to 1998, net of any recoveries, are as follows.

21 *ibid* 495.

22 [2000] 2 WLR 1007.

23 Not costs, as the Court did not allow any of the barristers to claim fees for the wasted costs hearing.

24 *The Times*, August 12 1999.

25 [1998] CLY 1016.

26 [1997] PNLR 489, CA.

27 [1999] COD 119.

28 *The Times*, July 26 1999. The case is reported as *R v Qadi* at [2000] PNLR 137, which records the wasted costs as £1,300.

29 [1996] FLR 41.

Year	Number of claims	Successful claims (judgment or settlement)	Average payment per successful claim	Number of claims where costs incurred	Average costs of those where incurred
1994	102	33	£5,855	25	£6,800
1995	85	23	£4,423	27	£5,663
1996	104	18	£5,866	23	£9,261
1997	99	22	£2,709	40	£5,497
1998	85	20	£5,434	12	£16,128

Thus over five years, BMIF have paid out in claims £568,748, at the cost of £848,868, which is £1.49 for each pound recovered.³⁰ The true position is surely very much worse, because in most cases the applicant fails to obtain any award, but will incur costs in attempting to do so. The small size of many of the claims is also noteworthy. Of the 475 applications, the wasted costs sought were less than £2,000 in 197 cases, and between £2,000 and £20,000 in a further 216.³¹ It should be emphasised that these figures are for the five years after the Court of Appeal in *Ridehalgh* had attempted to lay down guidelines to discourage applications which would be more expensive than the wasted costs which were in issue.

Judge-initiated applications

Compared with ordinary litigation, the wasted costs jurisdiction has a number of serious hurdles, which affect the applicant probably more than the respondent, quite apart from the disproportionate costs involved, and I will consider them in this and the following sections.

The first hurdle is that wasted costs jurisdiction can be out of the control of the litigants, as the Court itself can initiate a wasted costs application. The non-respondent party will almost inevitably be committed to the expenditure of costs which he did not choose and does not arise from any conceivable fault on his part.³² The Court of Appeal in *Ridehalgh* cautioned against judge-initiated enquiries in moderate terms:³³

In straightforward cases (such as failure to appear, lateness, negligence leading to an otherwise avoidable adjournment, gross repetition or extreme slowness) there is no reason why it should not do so. But save in the most obvious case, courts should in our view be slow to initiate the inquiry.

I have found only one reported example where it is entirely clear that judges initiated the enquiry, which may suggest that the courts have properly followed the indications in *Ridehalgh*. However, in very many reports it is just not clear whether

30 The figures may be distorted by the fact that costs applications or orders may not be reported to BMIF, particularly in small cases where there is £250 deductible.

31 There were also 49 cases where the costs sought were between £20,000 and £90,000, and the remaining 13 were over £90,000. Over the same five year period, claim payments were made in: 48 cases where the costs sought were less than £2,000, at an average per case of £930; 56 cases where the costs sought were between £2,000 and £20,000 at an average of £4,194; 11 cases where the costs sought were between £20,000 and £90,000 at an average of £21,779 per case; and one payment where more than £90,000 was sought which was settled at £49,750.

32 Given the BMIF figures set out in the next paragraph, one would expect that the costs greatly outweigh the likely recovery.

33 n 5 above, 238.

the Court or a litigant is responsible, and it must be highly likely that the Court did initiate the enquiry in a large proportion of them, particularly criminal and family cases. *Re G and ors (children)*,³⁴ where the Judge did launch the enquiry, had a mixed outcome and it is a salutary example which suggests that Courts should be very cautious before doing so. It concerned a case of negligence leading to an otherwise avoidable adjournment, and arguably fell squarely within what *Ridehalgh* permitted. However, after the Judge initiated the enquiry, he was surprised with the substantial submissions generated from what he thought was a simple point, and made a wasted costs order against only one of the four barristers and in respect of only one of the two charges. The time spent in the wasted costs hearing far outweighed the wasted costs.

A much clearer picture is given by BMIF's statistics between 1994 and 1998 inclusive. In that period, the number of judge-initiated enquiries were 4, 5, 10, 14, and 16. This is a little over 10 per cent of the total of wasted costs applications. One would expect that the greatest impact of the Woolf reforms on this jurisdiction would be to encourage judges to initiate wasted costs applications, given the emphasis on the court being proactive, and thus the figures since 1998 may well be very much worse. The outcome of such enquiries has been very poor. Of the 49 applications, wasted costs orders were made in only 6 cases, ignoring those quashed on appeal, a success rate of 12 per cent compared with 24 per cent for the wasted costs jurisdiction as a whole. BMIF paid £13,450 in wasted costs, with total defence costs of £287,078.³⁵ This is a cost of £21.34 for each pound recovered, compared with £1.49 for the wasted costs jurisdiction as a whole. This is wholly unacceptable.

The party in whose favour a wasted costs order might be made may well not wish a Court to initiate a wasted costs enquiry, because it may rightly feel that it is not economic to do so, but it can do little to prevent the Court embarking on a frolic of its own. But the greatest injustice is to the respondent lawyer. The same judge will both initiate the enquiry and come to a conclusion at the end of it. No other tribunal would be allowed to be prosecutor, witness and judge, and it is in principle wrong that our courts should be forced to act in a way which is arguably in breach of natural justice. It may also be in breach of Article 6 of the European Convention on Human Rights,³⁶ now enshrined in English law by the Human Rights Act 1998. These problems appear in a lesser way if the Court suggests the making of an enquiry. Even if the enquiry is wholly driven by the applicant, the judge may still be a witness. The law requires that the judge who dealt with the underlying action should also hear the wasted costs application.³⁷ The reason behind this rule is the laudable aim of saving costs,³⁸ and because the judge will be in the best position to assess the lawyers conduct,³⁹ but the result is that the hearing does not appear to be as impartial as it otherwise would be. We have thus ended up with a jurisdiction in which justice is administered less fairly than in normal litigation, but which in practice is far more expensive for the benefits gained. Appeals are discouraged,⁴⁰ which attempts to save costs at the price of

34 See text at n 22 above.

35 The other party to the litigation will normally also incur costs, so the true position is very much worse.

36 For an analogy, see the courts-martial case of *Findlay v UK* (1997) 23 EHRR 221.

37 *Bahai v Rashudun* [1985] 1 WLR 1337; *Re Freudiana Holdings Ltd*, *The Times*, December 4 1995; *Re Merc Property Ltd*, *The Times* May 19 1999.

38 *Bahai v Rashudun* [1985] 1 WLR 1337, 1346.

39 *Fryer v RICS* [2000] Lloyd's Rep PN 53, para 40.

40 *Wall v Lefever* [1998] 1 FCR 605, *Fryer v RICS* [2000] Lloyd's Rep PN 534 at para 39, both CA.

further potential injustice. If the wasted costs application is before the Court of Appeal, then it is difficult to see how the loser can even appeal in any sensible way.

Procedural difficulties

It would appear that a number of wasted costs applications fail, at least on appeal, as a result of a failure to follow the correct procedure. Thus in *Re Harry Jagdev & Co*⁴¹ the amount awarded was not specified, in *Livingstone v Frasso*⁴² the solicitors were not given an opportunity to show why the wasted costs order should not be made, in *Kilroy v Kilroy*⁴³ the judge had failed to identify the conduct which was improper, unreasonable or negligent, and had failed to assess the costs which had been wasted, and in *R v M (Wasted Costs Order)*⁴⁴ the judge had failed to address the issue of causation. One may think that this is the fault of the lawyers, rather than the inherent nature of the wasted costs jurisdiction, as the procedure is laid down in the cases, in particular in *Ridehalgh*,⁴⁵ and it is hardly a criticism of the necessary and proper procedural safeguards if advocates and judges do not adequately look up the law⁴⁶ before embarking on an enquiry. However, it is fair to say that 'making an application for a wasted costs order is not straightforward.'⁴⁷ Such difficulties can only encourage appeals on technical procedural grounds, although some of them fail.⁴⁸

Procedural defects can sometimes result from the fact that an apparent error comes to light, and an application is then immediately made for wasted costs. An example is *S v M*,⁴⁹ where the application was made near the start of a hearing after a vital allegation of forgery had been withdrawn, without any or much notice, due to lack of expert evidence. The order to show cause was made without a statement of what the lawyers had done wrong, and without satisfying the requirement that a strong *prima facie* case had been made before the respondent lawyers were called upon.

The wasted costs jurisdiction is summary. This should mitigate against the costs of the exercise being greater than the benefits, but it does not mean that the procedure is not without its complexities. Indeed, under the Civil Procedure Rules there are four stages (although the second and third may be combined): the applicant notifies the respondent of what he has done wrong and the costs sought; the court gives directions for summary disposal; the Court should then consider the application and be satisfied that there is material so that a wasted costs order is likely and the proceedings are justified; and there is finally further consideration of the matter after the respondent has put his case.⁵⁰

41 *The Times*, August 12 1999.

42 [1997] CLY 604.

43 [1997] PNLR 66.

44 [2000] PNLR 214.

45 n 5 above. There are, of course, genuinely difficult jurisdictional and procedural issues which do arise from time to time, compare *R v Camden London Borough, ex p Martin* [1997] 1 WLR 395 and *R v Immigration Appeal Tribunal, ex p Gulbamer Gulsen* [1997] COD 430.

46 Set out in the old and new White Books (except in the first edition of the latter), and the old and new Green Books.

47 Flenley and Leech, note 1 above, paragraph 11.3.

48 Eg *Re a Solicitor (Wasted Costs Order)* [1996] 1 FLR 40; *Woolwich Building Society v Fineberg* [1998] PNLR 216.

49 [1998] 3 FCR 665.

50 CPR part 48 practice direction section II.

There are four consequences of this summary procedure which should give serious concern to any litigant considering making an application. First, the application may be struck out if it becomes too complicated. As the Court of Appeal stated in *Ridehalgh*,

at the stage of initial application, when the court is invited to give the legal representative an opportunity to show cause ... [t]he costs of the inquiry as compared with the costs claimed will always be one relevant consideration ... judges may not infrequently decide that further proceedings are not likely to be justified.⁵¹

At least at such an early stage, the applicant should have spent comparatively little in costs. In complex cases, this sort of consideration may take place later, the outstanding example of this being *re Freudiana Holdings Limited*.⁵² At the stage the application was struck out the respondent's costs had reached £400,000. The applicant's lawyers were blamed for making the application too complicated, with many allegations including ones of dishonesty. It has been rightly pointed out, though, that 'it will be in the interests of those defending such applications to emphasise the extent to which a hearing will be long and cumbersome, dealing with many issues, and ultimately unlikely to produce a clear conclusion.'⁵³ Indeed, the effect of the law is to encourage respondent lawyers to take as many points as they can.

Secondly, even at the final stage, the court is not obliged to make an order, even if improper, unreasonable or negligent conduct and the causation of wasted costs is made out, although there have to be sustainable reasons to do so.⁵⁴ It must be unusual for an application to fail for this reason, but it is a further hazard the applicant must face. A rare example is *R v Secretary of State for the Home Office, ex p Tim Fat Wong*,⁵⁵ where the respondent acted unreasonably in failing to come off the record when legal aid was withdrawn, but no wasted costs order was made because otherwise the point which caused the applicant to win might never have come to light and an injustice would have been done.

Thirdly, one effect of the summary nature of the wasted costs procedure is that it is difficult to settle applications. There is a good reason why they should be disposed of swiftly, which is that everyone, including the judge, will still recall the details at the hearing, so less will need to be explained to the judge and there will be less room for dispute about what was said.⁵⁶ Furthermore, it is of course permissible for the parties to settle a wasted costs application,⁵⁷ and the Court of Appeal has approved a procedure for informing the court of such a settlement by short written statements.⁵⁸ But in practice compromise is much harder to achieve because of the summary nature of the procedure. While sheer lack of time is probably the most important factor, the absence of the normal provisions for mediation or part 36 offers or payments does not help. It is common experience in most actions which are fought that the costs are significant compared with what is at stake, and it is therefore extremely important that the parties should have every

⁵¹ n 5 above, 239.

⁵² *The Times* December 4 1995. The figure of £400,000 does not appear in this report, but see Flenley and Leech, n 1 above, at 11.8.

⁵³ Flenley and Leech, *ibid* at 11.9.

⁵⁴ see *Ridehalgh v Horsefield*, 5 above, 239.

⁵⁵ [1995] COD 331.

⁵⁶ See *In Re Merc Property Ltd*, *The Times* May 19 1999.

⁵⁷ Though presumably not an application initiated by the judge, see *Re a Barrister (Wasted Costs Order) (No 1 of 1991)* [1993] QB 293.

⁵⁸ *Manzanilla Ltd v Corton Property and Investments Ltd* [1997] 3 FCR 387.

opportunity to settle their dispute. Wasted costs applications, which are in any event more expensive for the sums at stake than normal professional negligence actions, do not have such effective opportunities.

Fourthly, a wasted costs order is only made in a clear case.⁵⁹ In practice this is likely to mean that it is harder to establish negligence in a wasted costs application rather than a normal action. A dispassionate consideration of the six appeals in *Ridehalgh*,⁶⁰ where none of the lawyers were considered to have acted unreasonably, improperly or negligently, suggests to me at any rate that the Court was being charitable to many of the respondent lawyers, and that the same result may not have been reached in an ordinary negligence action. The position is perhaps more complicated in the substantial body of cases where the respondent lawyers are criticised for bringing actions which are an abuse of process. *Ridehalgh*,⁶¹ suggests that it is not easy to define a case which is an abuse, rather than merely being hopeless, but it is not hard to say which is which. I do not think that experience bears this out: it is in practice quite difficult to determine whether a case is an abuse of process.⁶²

Privilege

In the rare case in which the lawyer's clients initiate an application for wasted costs, privilege will have been waived. In the more normal case, where the application is made by the opposing party, the privilege will not be for the lawyer to waive, and his client may often be unwilling to do so. The consequences were made clear in *Ridehalgh*:

So the respondent lawyers may find themselves at a grave disadvantage in defending their conduct of proceedings, unable to reveal what advice and warnings they gave, what instructions they received ... Judges who are invited to make or contemplate making a wasted costs order must make full allowance for the inability of respondent lawyers to tell the whole story. Where there is room for doubt, the respondent lawyers are entitled to the benefit of it. It is again only when, with all allowances made, a lawyer's conduct of proceedings is quite plainly unjustifiable that it can be appropriate to make a wasted costs order.⁶³

I do not think that these principles could be faulted. However, the difficulty caused by privilege fundamentally remains. The respondent lawyer may still suffer from not being able to explain what happened. On the other hand, the applicant is in an even less enviable position. He has no idea whether privilege will or will not be waived, and if waived what it might show; it is difficult to predict how the respondent lawyers might attempt to rely on privilege, impossible to know whether privileged material actually provides a defence, and in principle unpredictable how the Court will make allowances for privilege.

Privilege is relied on from time to time.⁶⁴ Two cases in particular illustrate the problems which may occur. In *Walter v Neville Eckley*,⁶⁵ the plaintiff sued the

⁵⁹ *Wall v Lefever* [1998] 1 PCR 604, 614C.

⁶⁰ n 5 above. The five cases are summarised in Jackson and Powell, n 1 above, para 4-58, 4-88, 4-107, 4-147 and 5-23.

⁶¹ n 5 above, 234.

⁶² See the cases collected in Jackson and Powell, n 1 above, at para 4-106 and 5-22, and also *Fletamentos v Effjohn*, Court of Appeal 10 December 1997, unreported, summarised in Flenley and Leech, n 1 above, para 11-31 and 11-32.

⁶³ n 5 above, 237.

⁶⁴ See eg *S v M* [1998] 3 PCR 664.

⁶⁵ [1997] BCC 331.

defendant liquidator for selling at an undervalue the premises belonging to his company which was in liquidation. An application for discovery was dismissed as misconceived, and another application was made against the plaintiff's solicitors to pay the wasted costs personally. The registrar made a wasted costs order despite the solicitors swearing that they had relied on Counsel's opinion, which was privileged. Sir Richard Scott V-C held that the registrar was wrong to make a costs order based on speculation that counsel's opinion would not have helped the solicitors. By the time of the appeal, however, privilege had been waived. The Judge held that the solicitors were entitled to rely on counsel's two opinions, although they were wrong in material respects. However, after the last of the opinions, further evidence came to light which should have caused the solicitors to reconsider the decision to make a discovery application, with or without counsel, and the wasted costs order was therefore affirmed on different grounds. Thus the registrar was right for the wrong reasons, as the perfectly proper reliance on privilege, which should have prevented an order being made, did not in fact relate to matters which would provide the solicitors with a defence.

In *Tolstoy-Miloslavsky v Aldington*,⁶⁶ the different views of the first instance judge and the Court of Appeal on the fact that there had been no waiver of privilege was a very significant factor in reaching opposite conclusions. The Court of Appeal made a wasted costs order against Count Tolstoy's solicitors for improperly running a hopeless case in very exceptional circumstances. The question in that case was whether privileged matter might provide any exculpation or suggest that there was additional material which had not been placed before the Court. In principle, surely, this might just be relevant. It is not perhaps entirely fanciful to suggest that a witness supported the claimants admittedly very weak case, but was unwilling to swear an affidavit to help resist the application to strike out.

The solution adopted in the new Civil Procedure Rules, part 48.7(3), was to permit the court to direct that privileged documents were to be disclosed. This solution rides roughshod over an important principle of justice, and, it may be thought, replaces one injustice with a greater one. In *General Mediterranean Holdings v Patel*,⁶⁷ Toulson J held that privilege was not a rule of evidence, but a fundamental principle of the substantive law, which could not be overridden save by express language or necessary implication, which he could not find in the Civil Procedure Act 1997. As a result, he struck out the new rule.

Contribution

This objection to the wasted costs jurisdiction is perhaps the least important because I have found no reported case where it appears to have been a real issue, and in practice there is a solution to what would appear to be a mostly theoretical difficulty. Suppose one legal representative, say the solicitor, is the respondent to a wasted costs order, and he seeks a contribution from another legal representative, say a barrister. How can he do this?

The normal method in an ordinary action would be to seek a contribution under the Civil Liability (Contribution) Act 1978. This might be done at the time of the wasted costs application. Alternatively the respondent solicitor has two years

⁶⁶ [1996] 1 WLR 736, 746, 748, and 749.

⁶⁷ [1999] 3 All ER 673.

thereafter before his action is statute-barred: this will mean he does not have to launch into contribution proceedings without time for considering whether it is cost-effective; on the other hand it contradicts the spirit of the wasted costs jurisdiction, which is in favour of a summary determination. However, it is very doubtful whether any such action is possible as a matter of law. The Act states in subsection 1(1):

any person liable in respect of and damage suffered by another person may recover contribution from any other person liable in respect of the same damage.

Subsection 6(1) provides that:

A person is liable in respect of any damage for the purposes of this Act if the person who suffered it ... is entitled to recover compensation from him in respect of that damage (whatever the legal basis of his liability whether tort, breach of contract, breach of trust or otherwise).

The applicant solicitor would not have been 'entitled' to recover compensation from the barrister, as the jurisdiction is discretionary. The wording of section 51(6) of the Supreme Court Act 1981 states that the 'the Court *may* disallow or (as the case may be) order the legal or other representative concerned to meet ... wasted costs'. There is also express reference to discretion in *Ridehalgh* and other cases. The word 'entitled' would appear to be important.⁶⁸

It can be argued that there is an alternative to a normal contribution claim. Pursuant to section 51(6) of the Supreme Court Act 1981,⁶⁹ the Court might make in the applicant solicitor's favour a further costs order by way of contribution or indemnity against the barrister, as such an order would be in respect of wasted costs "incurred by a party". In the unreported case of *Fletamentos v Effjohn*,⁷⁰ Simon Brown LJ thought this was not correct; Morrit LJ considered that it may be; but both opinions were brief and *obiter*. Certainly, any such argument faces the formidable difficulties that a wasted costs order cannot be made in favour of a non-party,⁷¹ save under the inherent jurisdiction,⁷² which of course requires the proof of serious misconduct.

There may be practical solutions to these difficulties. Two possibilities were raised by Counsel in *Fletamentos*, and are mentioned without comment by Simon Brown LJ. The court can initiate a wasted costs application against the second lawyer of its own motion. However, the Court would normally be interfering with the decision of the applicant that it was not worthwhile to seek costs from the second lawyer as well as the first. Alternatively, the applicant should be taken to have waived such part of their entitlement as would properly have been recoverable from counsel. This is even more objectionable: it in no way solves the problem; it is wholly contrary to normal common law rules in contract and tort; and it encourages the proliferation of parties and thus costs in this already costly jurisdiction.

In practice, it is not that uncommon for a Court to order that one legal representative pays x per cent and another y per cent of the wasted costs. For instance, in *R v Secretary of State of the Home Department, ex p Begum*,⁷³ the solicitors wrongly relied on Counsel's opinion in a judicial review application that

⁶⁸ See *Lister v Thomson* [1987] 1 WLR 1614.

⁶⁹ ie the statutory section concerning wasted costs orders.

⁷⁰ Court of Appeal, 10 December 1997.

⁷¹ *R v Camden London Borough Council ex parte Martin* [1997] 1 WLR 59.

⁷² *R v Immigration Appeal Tribunal ex p. Gulbamer Gulsen* [1997] COD 430.

⁷³ [1995] COD 176.

material letters did not have to be put before the Court. The barrister had to pay three-quarters of the wasted costs, and the solicitors a quarter.

An illogical solution

Perhaps the fundamental flaw of the Wasted Costs jurisdiction is that it is an attempt to remedy what have been two other defects in the legal system. First, there is the principle of advocate's immunity. There is a fundamental contradiction between this immunity, and permitting the recovery of wasted costs against an advocate for his conduct in court and what is intimately connected with it. There is a similar tension between denying any duty of care by a lawyer to his client's opponent (save in exceptional cases), and permitting the latter to recover wasted costs from the lawyer. It is no real justification to say that the duty is owed to the Court. Secondly, there is the unjust legal aid provision that neither the Legal Aid Board nor the legally aided litigant who loses will normally have to pay costs to his successful opponent, which has little more logic than saying that a legally aided defendant should not have to pay damages.

Without these two principles, the wasted costs jurisdiction would be invoked far less often.⁷⁴ Absent the protection of legal aid (now community legal service funding), the injured litigant could recover costs from his opponent, and through him, by way of assignment or bankruptcy if necessary, he could acquire a cause of action against his opponent's lawyers. Without immunity, the opponent could recover the costs he has had to pay to the injured litigant, and his own wasted costs, from his lawyers.

As it has existed up to now, the wasted costs jurisdiction is fundamentally an additional liability a lawyer will have to bear if he acts for a legally aided litigant. The Court in *Ridehalgh* was alive to this issue: 'It is incumbent on courts to which applications for wasted costs orders are made to bear prominently in mind the peculiar vulnerability of legal representatives acting for assisted persons'.⁷⁵ But the point remains that most applications are by one party against the legal advisers of his legally aided opponent, and that it is difficult to see why a party would ever make a wasted costs application against his opponent's lawyers unless the opponent was legally aided or impecunious.

Now advocates' immunity has been abolished, one part of this fatal mechanism will go. The government's preferred solution to the problems of legal aid is conditional fees. While this has received widespread opposition, one small incidental benefit should be to reduce the number of applications for wasted costs made by parties against their opponents' solicitors, because the litigant himself should have insurance to cover such costs. It is to be hoped that the courts will also curb their enthusiasm for a jurisdiction which has been shown by experience to be inherently defective.

74 There are a few cases where the losing party is impecunious, such as *Tolstoy-Miloslavsky v Aldington* [1996] 1 WLR 736.

75 n 5 above, 234. A point repeated in *Wall v Lefever* [1998] 1 PCR 605.

LEGISLATION

Democracy by Default: The Representation of the People Act 2000

*Heather Lardy**

Electoral law reform plays a significant part in the current government's programme of constitutional development. The devolution settlements, reform of the voting system for European Parliamentary elections, and the report of the Neill Committee on political party funding and election campaign expenditure have all necessitated consequential changes to the law governing elections.¹ Perhaps less visible, however, are the changes to the law governing parliamentary and local government elections which have been enacted in the Representation of the People Act 2000.² This statute implements the recommendations of the Howarth Working Party on Electoral Procedures, established by the government in early 1998 to conduct a broad review of electoral law and procedure, with particular emphasis on declining rates of participation at elections.³ The RPA 2000 seeks to encourage voter participation by modernising the system of electoral registration, by providing for postal ballots on demand, and by authorising the use of experimental methods of vote-casting and counting at local elections.⁴ The statute does not consolidate the law contained in the earlier Representation of the People Acts, instead making major amendments to the RPA 1983. This note comments on the main provisions of the Act, offering the view that its reforms, while generally welcome, may be undermined by their neglect of basic issues of principle concerning the enjoyment and exercise of the right to vote.

*University of Aberdeen. Thanks to Angus Campbell, Oonagh Gay and an anonymous referee for their comments on earlier drafts. Any remaining errors are mine.

- 1 Scotland Act 1998, ss 1–18; Government of Wales Act 1998, ss 1–15; Northern Ireland (Elections) Act 1998, ss 1–5; European Parliamentary Elections Act 1999. The proposals of the Committee on Standards in Public Life (Neill Committee) in its 5th report, *The Funding of Political Parties* (October 1998, Cm 4057) are implemented by the Political Parties, Elections and Referendums Bill 2000.
- 2 Hereafter 'the Act' or 'the 2000 Act' or 'RPA 2000'. 'RPA' is used throughout to denote a Representation of the People Act.
- 3 *Final Report of the Working Party on Electoral Procedures* Home Office October 1999. The Government clearly directed the Howarth Working Party to have regard to 'the Government's concerns that interest in the democratic process, as measured in part by participation in elections at all levels, has shown a steady decline over a number of years', para 1.1. The Select Committee on Home Affairs issued a report on electoral law during the deliberations of the Working Party, which was also focused on the need to increase voter turnout: *Electoral Law and Administration*, Select Committee on Home Affairs Fourth Report 1997–98.
- 4 The Act also contains more minor provisions which are not discussed. They concern service voters (s7) assistance for partially-sighted voters (s13) and free delivery of election addresses at Greater London Authority elections (s14). In addition, the Act amends the RPA 1983 to add a new offence of making false statements about a candidate's name or address in nomination papers (RPA 1983, s65A inserted by RPA 2000, Sched 6 para 5): this is to plug a gap left by the Registration of Political Parties Act 1998, which tackled only the problem of misleading party names or descriptions. (The 1998 Act will be superseded by Part II of the Political Parties, Elections and Referendums Bill 2000.) Forging the signature of an assentor to a nomination, or extracting it duplicitously is also made an offence by RPA 2000, Sched 6 para 5.

Entitlement to Vote

Like its predecessors, the RPA 2000 maintains the distinction between entitlement to vote and entitlement to electoral registration.⁵ The right to vote remains conditional on being registered to vote, on not being subject to any legal incapacity (eg conviction of a 'corrupt practice' offence), and on meeting the age and citizenship requirements. The right to be registered to vote is still determined by residence, in conjunction with the age, citizenship and incapacity rules. However, the 2000 Act makes certain changes to entitlement to registration which have the effect of amending indirectly the law governing the franchise, extending the right to vote to potentially significant numbers of people who previously experienced a *de facto* disenfranchisement due to the restrictive nature of the registration laws.

Electoral Registration

The RPA 2000 introduces a new system of voter registration, replacing a scheme which was undoubtedly in need of reform. Under the RPA 1983 a person's register entry was effectively fixed for sixteen months from the 10 October census date each year. The register took effect for one year from the following 16 February and during its currency no alterations could be made in response to voters changing address or seeking to register late. This inflexible system rendered the register notoriously inaccurate. More worryingly, it served as a potential barrier to the exercise of rights to electoral participation. Getting registered to vote (and particularly getting re-registered elsewhere) could prove to be a slow and cumbersome bureaucratic exercise, and may have acted as a deterrent to participation. Voters who changed address during the currency of a register, for example, were required to wait for up to sixteen months before becoming entitled to re-register at their new address. They remained entitled, in the meantime, to apply for an absent vote enabling them to cast a ballot in their old constituency, but they were nonetheless temporarily excluded from participation in electoral affairs in their new community. And for many, the further administrative hurdles involved in applying for an absent ballot no doubt provided a considerable disincentive to exercising their voting rights.

Under the new system of rolling registration instituted by the RPA 2000 a voter may be entered on the register at any point in the year by making an application to the local electoral registration officer to be included.⁶ Although this still requires compliance with bureaucratic formalities, it is preferable to the old system in that the voter is granted registration, and the right to vote, on the basis of her new place of residence, rather than on the rather artificial basis of her prior connection to another location.

5 Entitlement to vote is set out in RPA 1983, new ss1 and 2, substituted by RPA 2000, s1(1); entitlement to registration is in RPA 1983, s4, substituted by RPA 2000, new s1(2).

6 The idea of a rolling register has been around for some time, and was the subject of private member's Bills introduced by Harry Barnes MP in the 1993-94 and 1994-95 parliamentary sessions: see Representation of the People (Amendment) Bill HC Deb vol 237 col 809 15 February 1994 and Representation of the People (Amendment) (No 2) Bill HC Deb vol 265 col 309 1 November 1995. Political concern about low levels of voter registration was heightened by the role of the 'poll tax' in discouraging compliance with electoral registration procedures in the late 1980s and early 1990s.

Registration does not take effect immediately, but the delay between applying to be registered and the register entry taking effect should amount to no more than six weeks, and may be no more than two.⁷ If an election is pending, however, a revised register entry will not be effective for that election if it takes effect after the final day for the nomination of candidates.⁸ It is interesting that both the Home Affairs Select Committee and the Howarth Working Party had recommended an earlier cut-off point, of the day on which the election was announced.⁹ Both the Select Committee and the Working Party had taken the view that this would minimise the risk of abuse by, for example, individuals seeking registration in a constituency by establishing a temporary residence there after the election had been announced. There may be a (fairly minimal) risk of such abuse, but it is surely countered by the need for the system to respond to the registration requests of all those individuals who become motivated to check their inclusion on the register only after the announcement that an election is imminent. It might be argued that an even later final date would have been both feasible administratively and sufficiently safe from the risk of abuse.

The new system does not represent a total transfer of responsibility for initiating voter registration from the state to the individual. Instead, the scheme retains the annual canvass of voters conducted by electoral registration officers, and this is now to be done with reference to residence on 15 (not 10) October each year.¹⁰ An updated register is to be published on 1 December each year. Thereafter, registration officers may publish interim revised versions at intervals of their own choosing throughout the year, with each such version remaining in force, subject to amendments, until the issuance of the next interim list.¹¹ That will in turn be superseded by the next updated register issued on 1 December.

A further change instituted by the Act is that, in addition to the full electoral register, each local authority will now be required to prepare an edited register.¹² Voters will be able, under regulations to be made under the Act, to choose to withhold their names and addresses from the edited version of the register, which will be made available to companies and other bodies for commercial use. Each head of household will be required to mark opt-out boxes relating to each member of the household on the form distributed during the annual canvass; the names of those voters will then be omitted from the edited list. The full electoral register will continue to be available for public inspection at post offices and libraries, and will remain available to the political parties for electoral purposes. It will also be available for credit reference purposes, the government having conceded during the Bill's passage that this particular commercial use of the register might justifiably be continued in light of the importance attached by both credit companies and consumers to the widespread practice of granting consumer credit with reference to register entries. Other commercial organisations will, however, no longer be entitled to buy copies of the full register to use to generate mailing

7 RPA 2000, Sched 1, para 6 inserting RPA 1983, s13A(2), providing that the alteration will take effect from the first day of the month following the application, or 'if that day is less than 14 days after that time, on the first day of the month immediately following that month'.

8 RPA 2000, Sched 1, para 6, inserting RPA 1983, s13B. Alterations made to rectify minor clerical errors or to give effect to successful appeals against registration decisions may be effective after that date, provided they are made at least five days before the election: RPA 1983, s13B(1). The relevant date is determined in accordance with the parliamentary election rules RPA 1983, Sch 1 Part I.

9 The Home Affairs Committee Report, Section C para 42. Howarth Working Party Report, para 2.2.17.

10 RPA 1983, s10 substituted by RPA 2000, Sched 1 para 4.

11 RPA 1983, s13 substituted by RPA 2000, Sched 1 para 6.

12 RPA 2000, s 9 amending RPA 1983, Sched 2.

lists for unsolicited marketing campaigns. New regulations are to be made forbidding unauthorised copying or distribution of the full register, and specifying restrictions on the purposes for which the information in it may be used by those who have authorised access to it.¹³

This is a positive reform to the extent that it aims to reinforce the basic notion that the register is a public document which exists principally for electoral purposes, rather than for private or marketing ends. However, permitting the continued use of the register for credit reference does risk undermining its status as an officially sanctioned statement of democratic participants. No matter how convenient or efficient the register is as a basis for credit checks, there is no principled ground of distinction between that and other forms of commercial activity. If the register is to be withdrawn from some commercial purposes, it should be withdrawn from all. Credit reference agencies could use alternative sources to build the databases they require, in much the same way as other enterprises will now have to. Permitting the credit industry to continue to use the list grants it an unfair advantage over other commercial concerns, and does so by weakening the democratic credentials of the electoral register.

Residence

Residence remains the basic concept by reference to which entitlement to registration is determined under the RPA 2000, and it is to be assessed, as under the 1983 Act, having regard 'in particular, to the purpose and other circumstances, as well as to the fact, of his presence at, or absence from, the address on that date'.¹⁴ Except for electors in Northern Ireland, no particular period of residence is required to be established before entitlement to be registered to vote arises.¹⁵ Dual (or multiple) registration will remain possible for voters with two (or more) residences.¹⁶ This is so despite calls from within Parliament for a revision of this

13 RPA 1983, Sched 2 paras 10B & 11, amended by RPA 2000, s 9. The sale of the register raises issues of privacy and data protection, discussion of which is beyond the scope of this article. See eg Howarth Working Party Report, paras 2.4.7–2.4.25 & Annex 1 (extract of Data Protection Commissioner's submission to Working Party). The edited register reform is intended to ensure that the law complies with the Data Protection Act 1998 by providing that information supplied for one purpose is not used for another without the consent of the data subject.

14 RPA 1983, s5 as substituted by RPA 2000, s3.

15 Citizens living in Northern Ireland must have been resident there for three months prior to the date of registration (this rule is retained from RPA 1983, s1(2), see RPA 1983, new s4(2) substituted by RPA 2000, s1(2)). No corresponding duration-based residence test applies in other parts of the United Kingdom. The justification for the rule is to prevent citizens from the Irish Republic moving across the border just before an election, and registering to vote in marginal seats. That apart, the rule arguably discriminates in the terms on which electors living in Northern Ireland enjoy the right to vote, and might therefore be challengeable under the Human Rights Act 1998. Such an argument might be founded on Protocol 1, Art. 3, ECHR, which protects the right to vote and Art. 14, which protects, *inter alia*, against discrimination on grounds of national origin and of association with a national minority. It might be claimed that the three-month rule undermines the right to vote of those electors moving to live in Northern Ireland by reason of their association with (ie living beside) the national minority comprising residents of Northern Ireland. Alternatively, an argument might be made on the basis of Protocol 1–3 alone, claiming that the rule undermines the right to vote which that provision protects. See n 24 below.

16 Neither the 1983 Act nor its predecessors made express provision permitting or prohibiting dual registration, but the practice was sanctioned by the Court of Appeal in *Fox v Stark* [1970] 2 QB 463, which permitted students to register at both their halls of residence and home addresses. The Scottish courts, while recognising the possibility of dual registration, have denied it to voters who spend relatively little time at their second (holiday) home: *Ferris v Wallace* 1936 SC 561; *Scott v Phillips* 1974 SLT 32; cf *Dumble v ERO for the Borders* 1980 SLT (Sh Ct) 80.

rule.¹⁷ Objections to multiple registration tend to be based on the extra choice which it confers on the voters concerned.¹⁸ They may choose in which constituency to cast their ballots and thereby enhance the relative effectiveness of their votes by selecting the more marginal seat. This increased voting power is further boosted by the fact that those with multiple registration may also vote at local government elections at all their places of registration, and even, according to the terms of the RPA 1983, at by-elections in both (or all) seats at which they are registered.¹⁹ Some parliamentary supporters of multiple registration sought to invoke the slogan 'no taxation without representation', arguing that anyone who pays council tax for a property, even if it is not a principal residence, is thereby entitled to participate in local government elections.²⁰ However, many taxpayers who are lawfully settled in the UK are routinely excluded entirely from the franchise (both local and national) because they do not possess the requisite citizenship. Paying tax does not generate a basic entitlement to participate in elections, nor does it determine the extent of any electoral rights which are enjoyed.²¹

It is unfortunate that the opportunity was not taken to end multiple registration in this Act, particularly when an amendment to that effect was proposed during the Bill's parliamentary passage.²² Dispensing with dual registration would not cause any great hardship to those with more than one residence. The Act's new provision of postal votes on demand, coupled with the more flexible rolling registration procedure, would make it possible for such voters to shift their registration between their homes, or to seek a postal ballot if away from the home at which they were presently registered. Some possibility of manipulation of the system might remain, given that the rolling register would permit a voter to transfer her registration from one constituency to another after the point at which a general election were announced, but this might be tempered by the application of the criteria for determining residence in a manner which would deny registration to those who spent relatively little time at their second address.²³

More fundamentally, the practice of multiple registration is plainly discriminatory, and may be challengeable under the Human Rights Act 1998.

17 The Home Affairs Select Committee had suggested that voters with two (or more) residences should nominate their main residence when registering to vote and would have to vote there at national elections, *Electoral Law and Administration*, paras 119–121. The Working Party thought that this idea was neither practical nor desirable, finding that it would be complex to administer and might serve to discourage voting. Instead, the Working Party endorsed the continuation of multiple registration: 'in the absence of any clear case law, it would be unreasonable to deprive people properly registered in two places of the right to have a voice in electing a candidate to represent their interests on each of the local councils where they live' Working Party Report para 2.5.4.

18 See eg HC Deb vol 341 cols 340–365 15 December 1999.

19 See RPA 1983, new s1(2) (RPA 2000, s1), which restates the rule that no person may vote more than once in the same constituency at any parliamentary election or in more than one constituency at a general election. See also RPA 1983, s 61(2), rendering it an offence to vote in more than one constituency at a general election, or in more than one electoral area at a local government election; it is also an offence to vote more than once in the same constituency at any Parliamentary election, or more than once in the same electoral area at any local government election. Voting in two concurrent by-elections is not an offence; nor is voting in two separate local authority elections held on the same day.

20 See HC Deb vol 341 cols 341–347 15 December 1999.

21 Further, those with more than one home pay the full council tax rate only in respect of the 'main' residence, an abated rate applies to any other properties which are owned or occupied: Local Government Finance Act 1992, s11.

22 See HC Deb vol 341 col 340 15 December 1999.

23 See eg *Scott v Phillips* 1974 SLT 32 (holiday home deemed not to constitute a residence for purposes of electoral registration).

Protocol 1, Article 3 protects electoral rights, and is included in those Convention rights given domestic effect in UK law by the 1998 Act.²⁴ Also included is Article 14, which provides that there shall be no discrimination as to the terms on which the rights set out in the Convention (including the right to vote protected by Protocol 1–3) are enjoyed. Article 14 makes specific mention of property as a ground on which discrimination is prohibited. It might be argued that to permit those with multiple residences to participate in more elections, and on terms which confer on them a wider set of political choices, amounts to discrimination on grounds of ownership or occupation of property against those who are unable to avail themselves of the multiple registration facility. The right to vote of those discriminated against is diminished by the reduced opportunities for electoral participation open to them. An alternative or supplementary argument might be made solely on the basis of Protocol 1–3, claiming that the guarantee of free and fair elections contained in that provision is undermined by a law which confers unequal electoral opportunities on the basis of the number of residences each voter occupies.

Homelessness

For those who neither own nor occupy any property, the RPA 2000 brings a long overdue reform. Although homeless individuals were never expressly denied the right to vote by the terms of the preceding statutes, the requirement that registration be by virtue of residence operated in effect to exclude most of them. Some homeless people did manage, with considerable effort and perseverance, to become registered to vote, but they achieved this only because they managed to bring themselves within the law's residence rule. Only in those cases in which the homeless person could establish some sort of residence, such as a hostel, did registration tend to be deemed possible.²⁵ Others, who sought unsuccessfully to register, commonly ran into difficulties with electoral registration officers unsure whether the legislation permitted them properly to register a person without a home address. The residence rule combined with registration practices to bring about the virtual disenfranchisement of the many homeless people who did not enjoy the benefits of even a temporary address. The RPA 2000 addresses this problem by introducing an innovative rule permitting a potential elector to demonstrate 'notional residence' by means of a declaration of local connection.²⁶ A homeless person may make this declaration, stating his name, an address to which mail may

24 Protocol 1–3 provides that 'The High Contracting Parties undertake to hold free elections at reasonable intervals by secret ballot, under conditions which will ensure the free expression of the opinion of the people in the choice of the legislature.' The European Court of Human Rights has interpreted this as providing individuals with an enforceable right to vote, *Mathieu-Mohin and Clerfayt Case*, 2 March 1987, Series A, No. 113. See also *Matthews v UK* (1999) 28 EHRR 361.

25 See *Lippiatt v Electoral Registration Officer for Penwith District Council*, 21 March 1996 unreported, in which the county court upheld the claim of a homeless man to be registered using a day centre as an address. See *The Guardian* 30 August 1996, 'Builder without a home to stand as MP'. The Home Office subsequently issued guidance to electoral registration officers in England in Wales stating that the *Lippiatt* case was not to be taken as a precedent: Home Office Constitutional Unit guidance to electoral registration officers, 29 July 1996.

26 RPA 1983, s7B inserted by RPA 2000, s 6. The provision is new, but the idea is not. The scheme follows closely that advocated by Harry Barnes MP in his Representation of the People (Amendment) Bill 1997, which proposed a procedure involving 'declarations of locality', see HC Deb vol 300 cols 908–909 12 November 1997. Under the Act's provisions, a homeless person who has some form of temporary accommodation may still seek to register by virtue of residence, see RPA 2000, s 3 substituting RPA 1983, new s 5(2) 'a person ... staying at any place otherwise than on a permanent basis ... may in all the circumstances be taken to be ... resident there if he has no home elsewhere'.

be sent (or stating his willingness to collect such correspondence periodically from the electoral registration officer's office), and giving in lieu of an address 'the address of, or which is nearest to, a place in the UK where he commonly spends a substantial part of his time (whether during the day or at night)'.²⁷ This place might be a shop doorway, or a park. Registration by means of a declaration of local connection is valid for 12 months, but will be superseded by any subsequent registration by virtue of residence.²⁸

The declaration of local connection procedure is a significant development because it is the first time that the law has granted explicit recognition to a criterion other than residence at an address as a determinant of entitlement to voting rights. Prior to the 2000 Act, the residence rule, which was intended to function as an approximate indicator of local connection, served also in effect as a virtual property requirement, having the unintended consequence of barring those who neither owned nor occupied a property from electoral registration. The effect of maintaining residence at an address as the exclusive route to registration was thus to keep alive the ghost of the property-based franchise, formally removed from the law in 1918.²⁹ By demanding that voters either own or occupy premises at a recognisable (postal) address, the law in effect required that they had access to the financial resources or support necessary to make that possible. The 2000 Act removes those vestigial traces of the property qualification by providing for the first time an alternative method of registration which does not depend on the voter having access to the financial means necessary to enjoy the ownership or occupation of property. Apart from its potential benefit to individual electors, this reform is welcome because it strengthens the distinction between what might be termed the substantive and procedural notions of residence in relation to electoral registration. The substantive notion relates to the requirement that each voter have a residence (ownership or occupation of an identifiable property) in order to register. The procedural notion pertains to the allocation of voters to constituencies on the basis of their place of residence and to the calculation of constituency electorates based on the number of residents registered to vote there. Prior to the 2000 Act, the law blurred this distinction by failing to recognise the procedural residence of anyone without a substantive residence. This left those without a home to contend with the vagaries of a system largely unsympathetic to their efforts to establish their status as voters in spite their lack of a substantive residence. The new provision for local connection declarations marks an important shift towards a more procedural conception of residence, one which encompasses those who were unfairly excluded by the unqualified operation of the substantive residence rule, and which grants them for the first time a protected right to register to vote.

The Act qualifies the new right it confers, however, by limiting the possibility of making a declaration of local connection during the run-up to Parliamentary, Scottish Parliamentary and Welsh Assembly by-elections. New section S7B (6) of RPA 1983 provides that declarations made in the period from the date when the vacancy occurs until the final day for the nomination of candidates must contain an additional statement to the effect that 'during the period of three months

27 RPA 1983, s7B(4)(b), inserted by RPA 2000, s 6.

28 RPA 1983, s7C(2), inserted by RPA 2000, s 6; RPA 1983, s7B(2)(c) states that the declaration procedure may be used by a person who is not 'resident' in terms of the Act at any address in the UK.

29 RPA 1918. See D. Butler, *The Electoral System in Britain Since 1918* (2nd ed, Oxford: Oxford University Press, 1963) ch1; R. Blackburn, *The Electoral System in Britain* (London: MacMillan, 1995) ch 3.

ending on the date of the declaration, the declarant has commonly been spending a substantial part of his time (whether during the day or at night) at, or near, the required address.' The reason for this rule is to avert the perceived risk of a mobile homeless population flooding into a constituency in which a vacancy has arisen, registering by means of a declaration of local connection, and swaying the vote. This fairly unlikely prospect caused some disquiet during the parliamentary debates on the provision.³⁰ The real objection ought instead to be to the potential discrimination which this rule sets up. Under the new system of rolling registration, a propertied voter moving to a new constituency and seeking to register during the pre-election period would not be required to comply with any three month residence rule.³¹ A homeless person would, however, have to establish a local connection dating back three months. This is unfortunate because the announcement of the election is the very event which is likely to do most to motivate potential voters to seek to register. Yet from that point on, the homeless person faces a greater hurdle than others. Also, the Act does not permit a homeless person to make more than one declaration of local connection, thereby precluding the double (or multiple) registration which is permitted in the case of those with two (or more) homes.³² These discriminations in the terms on which the right to vote is offered to the homeless and to the propertied might be regarded as sufficiently serious to constitute a breach of the guarantee of non-discrimination in the enjoyment of electoral rights provided by Article 14 and Protocol 1-3, ECHR. Arguments similar to those suggested in relation to the issue of multiple registration generally might be made in a Human Rights Act challenge: limiting the electoral rights of homeless persons (by prohibiting multiple registrations and by restricting registration prior to a by-election) is an interference with their right to vote (Protocol 1-3) on the basis of 'property ... or other status' (Article 14), and thus violates their rights under both the above provisions of the Convention. Alternatively, an argument might again be made solely on the basis of Protocol 1-3, proposing that the court read the guarantee of free and fair elections contained in that provision as a prohibition on laws prescribing 'unfair' registration practices such as those which limit the electoral opportunities of homeless persons.

Given that the very purpose of the 2000 Act provisions is to provide homeless persons with means of access to electoral registration on terms which do not discriminate against them, this reduced level of political rights is unfortunate. Further, an argument based on the Human Rights Act might be strengthened by the fact that the 2000 Act purports merely to extend the electoral registration opportunities of homeless persons by building on an enfranchisement presumed already to have been established. That is, the law prior to the 2000 Act is not regarded as having denied homeless persons the right to vote, only as having hampered their registration as voters. Assuming this to be so, the registration procedures made available to such voters should surely be the same in extent as those available to other groups. If homeless persons have in fact always enjoyed the same (theoretical) right to vote as others, then that right should now be matched

30 See HC Deb vol 342 cols 930-937 19 January 2000; HC Deb vol 342 cols 1016-1018 20 January 2000; see also HL Deb vol 609 cols 963-970 14 February 2000, debating an (unsuccessful) amendment requiring (only) homeless voters to produce identification at the polling station.

31 Unless the new constituency was in Northern Ireland and the old one was not: RPA 1983, new s4(2), substituted by RPA 2000, s1(2).

32 RPA 1983, s7B(b) & s7C(2)(c), as inserted by RPA 2000, s6. This rule also appears to have been motivated by concerns about possible abuse of the system.

by registration procedures which correspond in terms of convenience and extent with those enjoyed by their more comfortably accommodated fellow citizens. To offer them more restricted registration procedures is surely to undermine their electoral rights, and to do so for reasons related to their status as homeless persons.

Further, the assumption that homeless persons *were* in fact enfranchised prior to the 2000 Act is itself questionable. It was simply presumed throughout the parliamentary debates on the Act that they were, and the same belief permeates also the reports of the Working Party and of the Home Affairs Select Committee. However, the long standing exclusion of homeless people from the electorate has not been due purely to administrative difficulties in the procedure of registration, but also to assumptions about *entitlement* to participate inherent in the laws which regulate voter registration. A rule which requires voters to have a fixed address conditions the enjoyment of the right to vote on the possession of a particular sense of social connectedness, crudely measured in this instance by the nature of the accommodation the person enjoys. There is, as noted above, an important difference between pure procedural formalities (such as complying with a deadline for registration) and substantive registration requirements predicated on the possession of a particular characteristic (such as having a residence). The declaration of local connection procedure is to be applauded insofar as it represents a step towards making the process of electoral registration into a purely procedural formality. An honest assessment of the reform should acknowledge also that its effect is not merely to enhance an existing (and largely theoretical) entitlement but to create a brand new right to vote for a vulnerable and socially excluded group. Even if the new rights remain largely unused, this law will remain important as a formal acknowledgement that persons who are homeless are members of the political community, entitled to the share the benefits of the right to vote with their propertied neighbours.

Citizenship

Voting may be regarded as a right attached to citizenship or as a right (not necessarily attached to the possession of state citizenship) to be granted to all resident individuals as a means of protecting their interests.³³ The UK combines elements of both those approaches in its electoral law, requiring that voters at Parliamentary elections be British, Commonwealth or Irish citizens:³⁴ the electorate thus extends beyond British citizens, but not so wide as to include all overseas citizens resident here. Further, the RPA 2000 has modified this rule by restricting entitlement to voter registration to 'qualifying Commonwealth citizens.' The effect of this is to exclude from the register any Commonwealth citizens who require but have not yet been granted leave to enter or to remain in the UK.³⁵

33 See H. Lardy, 'Citizenship and the Right to Vote', (1997) 17 OJLS 75.

34 RPA 1983, s1(1)(c) as amended by RPA 2000, s1. Citizens of the European Union resident in the UK may vote in local, European Parliamentary, Scottish Parliamentary and Welsh and Northern Ireland Assembly elections, RPA 1983, s2(1)(c), as amended by RPA 2000, s1; European Parliamentary Elections Act 1978, s3C(5); Scotland Act 1998, s11(1); Government of Wales Act 1998, s10(1); Northern Ireland (Elections) Act 1998, s2(2).

35 RPA 1983, s4(6) substituted by RPA 2000, s1(2). This provision was added as a Lords' amendment, see HL Deb vol 610 cols 456-457 29 February 2000, apparently in response to concerns that the political composition of those constituencies housing high numbers of refugees might be affected by the votes of any Commonwealth citizens of voting age. The voting rights of Irish citizens remain unchanged.

Although the Act attempts to preserve the general principle of their eventual entitlement by retaining an unqualified statement of the rights of *any* 'Commonwealth citizen' to vote in the new section 1 of the RPA 1983, the refinement that only those who meet the 'qualifying' criterion may *register* to vote represents a clear shift away from automatic enfranchisement on the basis of a preferred connection between the voter's country of origin and the UK.

Calls were made during the Bill's parliamentary passage for the extension of the franchise to all overseas citizens resident in the UK, sparking a debate about the relative virtues of (state) residence and citizenship as determinants of the right to vote.³⁶ This exchange reflects a basic disagreement about the nature of the principle according to which the franchise should be distributed. It is in effect a dispute about the character of what might be termed the criterion of membership, that is, the collection of rules governing admission to the electorate. The proponents of enfranchising all who live within the state espouse a model of democracy according to which anyone who is affected by the laws made within the territory should have a voice in the selection of legislators, with which to assert and protect her private interests.³⁷ According to this view, membership is important principally because political rights are attached to it: non-members lack full political rights with which to assert their claims and defend their interests. By contrast, those who advocate the use of a citizenship-based criterion tend to regard voting as essentially a duty of citizenship, emphasising the ideas of loyalty and allegiance to the state.³⁸ This conception of membership regards the status of being a citizen-participant as intrinsically valuable, apart from (or in addition to) the value of the voting privileges which it confers: being a voter ranks with being a citizen and a taxpayer among the badges of civic virtue which the state awards.³⁹

In addition to the basic membership test, electoral law employs two further means of delimiting entitlement to vote. These are the criteria of competence and of virtue. Thus, for example, the criterion of competence is used to deny voting rights to those with a presumed incapacity to exercise political choice in a meaningful way; this is used to disenfranchise children. The criterion of virtue is used to disenfranchise those whose standards of behaviour lapse below the minimum expected of virtuous citizens; this underlies laws which strip some convicted criminals of the right to vote. The criteria of membership, of virtue and of competence are not simply procedural devices for drawing some sort of line around the electorate. They are instead mechanisms which convert (often ill-

36 HC Deb vol 341 cols 293–305 15 December 1999. There was also a debate about the extent of voting rights for UK citizens resident overseas, see HC Deb vol 342 cols 306–338 12 January 2000. Those rights stand to be limited by the Political Parties, Elections and Referendums Bill 2000, cl 41 which reduces the period of overseas residence during which such rights are available from 20 to 15 years.

37 See eg Harry Barnes MP: 'I seek to enfranchise people who are dependent on decisions made by elected representatives on laws, basic services, environmental protection, health and safety provision, taxation and economic prosperity. They affect the lives of anyone who has settled in this country but not taken British citizenship.' HC Deb vol 341 col 293 15 December 1999.

38 See eg Robert MacLennan MP: 'The link between citizen and state is a reciprocal one: the state has the duty to look after its citizens and the citizen must exercise his or her duty to be interested in how that service is provided', HC Deb vol 341 col 295 15 December 1999.

39 Pushed to their limits, neither of these principles can alone offer a comprehensive and coherent scheme for the distribution of voting rights. According to the former there would be no justification for excluding young children from the franchise; they are, after all, affected by the laws of the state. As regards the latter, the principle would deny the vote to some long-term residents who may have a deep allegiance to the state, yet lack the formal status of citizen. Each of these principles is therefore modified in operation by the application of other criteria.

considered and barely articulated) value judgments about who is fit to vote into suffrage laws. They can be observed at work in the operation of the provisions of the RPA 2000 which pertain to the franchise. The disenfranchisement of those Commonwealth citizens who have not complied with immigration requirements represents an expression both of the membership and the virtue criteria: only Commonwealth citizens who have complied with the relevant entry rules are deemed to possess the virtue demanded of those who seek to take advantage of the relaxed membership rules which grant some non-UK citizens access to voting rights. Similarly, the new registration rights of homeless persons represent a revised expression of the membership criterion, re-affirming their place as members of the political community. The criteria are evident also in the provisions outlined in the following discussion.

Remand Prisoners

The RPA 2000 enfranchises prisoners on remand, who were previously entitled to be included in the electoral register, but only if resident at their home address on the relevant date. They were not permitted to use the prison as an address for the purposes of voter registration. The Act tackles this problem by directing registration officers to regard such a prisoner as resident at the place of his detention 'if the length of the period which he is likely to spend at that place is sufficient for him to be regarded as being resident there for the purposes of electoral registration.'⁴⁰ The rationale behind this reform is that remand prisoners have not yet been convicted of any offence, and so should retain the right to vote pending the outcome of the trial. By contrast, section 2 of the RPA 2000 explicitly disenfranchises offenders who are detained in a psychiatric hospital following conviction.⁴¹

These apparently minor reforms conceal a basic question about the nature of entitlement to vote in UK law. The 2000 Act enfranchises only those prisoners who have not been convicted of the offence in relation to which they are detained. Convicted prisoners remain disenfranchised.⁴² There is, however, a strong case to be made for changing the law to permit convicted prisoners to vote. This issue was raised during the parliamentary debates on the Bill, with advocates of enfranchising prisoners arguing that as prisoners are affected by the laws of the state, they should have a voice in selecting legislators.⁴³ Indeed, prisoners arguably have a greater stake than many others in the changing shape of the law, particularly as regards issues of criminal justice. And this disenfranchisement might form the basis of a challenge under the Human Rights Act 1998: the withdrawal of the right to vote might be argued to be an unjustifiable interference with the electoral rights protected by Protocol 1–3 of the Convention. The Home Office minister did offer the prophylactic pronouncement during the Bill's Commons' passage that 'the disenfranchisement of prisoners is consistent with the jurisprudence of the

40 RPA 1983, s7A inserted by RPA 2000, s5. Note that such prisoners may continue to be registered by virtue of residence at a home address, or by making a local connection declaration, where appropriate: s7A(5). Whichever means is used, the registration runs for 12 months, or until another register entry takes effect: RPA 1983, ss7A(3) & 7C(2) added by RPA 2000, s6.

41 RPA 2000, s2 inserts RPA 1983, s3A which sets out the various statutory bases for the relevant detention eg Mental Health Act 1983; Mental Health (Scotland) Act 1984; Criminal Procedure (Insanity) Act 1964; Criminal Procedure (Scotland) Act 1995. This provision is a corollary to RPA 2000, s4 which enfranchises other categories of patients in psychiatric hospitals.

42 RPA 1983, s3.

43 HC Deb vol 342 cols 342–343 12 January 2000.

Convention.⁴⁴ It is true that the European Court of Human Rights has held that states have a wide margin of appreciation in relation to Protocol 1-3, but that does not of itself provide assurance that the courts of the United Kingdom will take a similar view of the matter.⁴⁵ It might reasonably be argued that evolving norms governing the grant of voting rights, and changing social understandings of the nature and aims of penal policy, merit a divergent approach to the interpretation of Protocol 1-3. Further, underlying arguments about the human rights issues there is a tension here between the two variants of the criterion of membership mentioned above. Those who argue for the extension of voting rights to prisoners do so mainly on the ground that convicts remain affected by law and law reform during their incarceration, and so should have a vote to cast in choosing lawmakers. Those who defend their disenfranchisement are basically arguing that the prisoners' criminal activity has placed them (either temporarily or permanently) outside the boundaries of the political community: their lapse in civic virtue causes them to experience a loss of membership, and consequent disenfranchisement. This is so because voting is, according to this view, a duty of the good citizen, and not, as the first view would hold, a basic right of all affected by the laws made by elected legislators. By transgressing, the criminal has displayed a lack of the virtue demanded of voters. It is this, rather than the fact that the prisoner still has interests to protect and is competent to exercise political choice, which governs the decision to disenfranchise.

The problem with this argument is that initial decisions about enfranchisement are not generally predicated on the notion of virtue: the law distributes the right to vote without reference to the idea, invoking it later only as a means of withdrawing the franchise from certain transgressors. While it might be claimed that all voters are required to maintain the same standards of virtue in order to retain their right to vote, this does not actually hold. Two individuals might commit (separately) the same type of offence; one might be sentenced to a prison term while the other might escape with a fine. Why should only the former lose the right to vote? Her lapse in virtue is surely no greater than the other's and yet she suffers the additional penalty of disenfranchisement. A Human Rights Act challenge to this law should seek to convince the court of the inappropriateness of employing the criterion of virtue in this random and selective manner, by emphasising instead the argument that prisoners are in fact displaced members of a political community in whose affairs they still have an interest and should have a voice.

Psychiatric In-Patients

Under the rules contained in the RPA 1983, voluntary psychiatric patients (ie those not detained in hospital under the relevant mental health legislation)⁴⁶ were entitled to be registered at the address at which they would have been living if not in hospital. An attested declaration had to be made by the patient (without any assistance) stating that she was a voluntary patient, qualified to vote and giving the address of the hospital and an address where she would be resident if not in hospital.⁴⁷ This procedure served, in effect, to assess the ability of the patient to understand the process of registration and voting: those capable of making the declaration unaided were deemed competent to vote. Registration achieved by this

44 George Howarth *ibid* col 343.

45 *H v Netherlands* (1983) 33 DR 242.

46 Mental Health Act 1983; Mental Health (Scotland) Act 1984.

47 RPA 1983, s7.

means was at the patient's home address, not at the hospital. Detained patients were not entitled to be regarded as resident at the hospital, and could not therefore be registered to vote there.⁴⁸ The fact of such patients' detention raised a presumption that their illness was sufficiently severe to disqualify them from registration and voting rights. Some detained patients might have been able to remain registered at another address, but for many this would have been impossible in light of the length of their hospital stay. The law thus effectively disenfranchised many detained patients, and made registration more burdensome for voluntary patients by requiring an attested declaration. The RPA 2000 sweeps away those restrictions, introducing a new rule whereby both the voluntary and the detained patient alike may be regarded as resident at the hospital 'if the length of the period which he is likely to spend at the hospital is sufficient for him to be regarded as being resident there.'⁴⁹

This welcome change represents a developing public understanding that mental illnesses do not necessarily disable individuals from exercising political choice. That said, it is not clear if the traditional exclusion of the mentally ill has ended entirely with the RPA 2000. There remains a long-standing common law restriction on voting with a mental disorder, the status and extent of which is uncertain. This rule, endorsed as extant by some current constitutional texts, permits officials to refuse to register, or to issue a ballot paper to, a person who appears to have a mental incapacity (eg illness or learning difficulty) which is sufficiently serious to deprive the person of the ability to understand the meaning of casting a vote.⁵⁰ The possibility thus exists that a psychiatric in-patient who has gained the presumptive right to register from the RPA 2000 might nevertheless find that entitlement being challenged by an electoral registration officer employing the common law rule. This is so because it appears that the common law may not have been entirely displaced by the new statutory rules: the Act permits a patient to use a psychiatric hospital as an address for the purpose of registration, but it does not confer an explicit entitlement on the patient to register or to vote regardless of her mental state. Further, the parliamentary election rules permit the presiding officer at the poll to put to a voter whose entitlement to vote is uncertain, two questions requiring her to confirm her register entry and that she has not voted already. These are the only questions that may be put. The officer has a duty to deny a ballot paper to someone unable to answer those questions satisfactorily.⁵¹ This test may be invoked to assess the capacity of a voter suspected of having some mental

48 RPA 1983, s7(1). The changes in the RPA 1983 consolidated reforms contained in the Mental Health (Amendment) Act 1982, s62, which had amended the RPA 1949, s49(3) to allow voluntary patients to register.

49 RPA 1983, s7(2), as amended by RPA 2000, s4. Alternatively, the person may, where appropriate, register by virtue of a residence other than the hospital or by way of a declaration of local connection, RPA 1983, s7(5). Registration by any of those methods will remain valid for 12 months. There is no obligation on returning officers to conduct a canvass of psychiatric hospitals within their area, an omission which might diminish the effectiveness of this reform. A Lords amendment requiring such a canvass was unsuccessful, see HL Deb vol 609 cols 974–976 14 February 2000.

50 See Little, *Schofield's Election Law* (London: Shaw & Sons, 1984) para 4.04.04. The only three cases cited there in illustration of the disqualification all date from the late 18th and early 19th centuries. The rule appears to apply to mental illness and to other forms of mental incapacity (eg learning difficulties; temporary impairment due to alcohol or drug use). In older texts it is crudely described as prohibiting 'idiots' from voting, while permitting 'lunatics' to do so during lucid intervals, see eg A.N. Schofield, *Parliamentary Elections* (London: Shaw & Sons, 1950), 20. Recent references to the rule include I. Loveland, *Constitutional Law: A Critical Introduction* (London: Butterworths, 2000), 196; A.W. Bradley & K.O. Ewing, *Constitutional and Administrative Law* (12th ed, London: Longman, 1997), 171; C. Munro, *Studies in Constitutional Law* (2nd ed, London: Butterworths, 1999), 95; R. Blackburn, *The Electoral System in Britain* (London: MacMillan, 1995), 83.

51 rule 35, parliamentary election rules. RPA 1983, Sched 1.

impairment likely to interfere with her understanding of the meaning and process of voting.⁵² This procedure has thus displaced the common law rule to the extent that the latter might have permitted officials to put a range of questions to the voter with a view to establishing fitness to vote. However, it is uncertain whether the common law would still permit an official to deny a vote to someone who gives satisfactory answers to the two questions above, but who nonetheless appears disorientated or deluded at the poll. A similar uncertainty arises about a possible refusal of registration based on the electoral registration officer's possession of direct knowledge or information about the applicant's mental state.

It is at least arguable that the parliamentary election rule setting up the statutory question procedure should be read as having replaced the common law rule, in that it provides a substitute test for competence based on the ability to answer two simple questions. No other questions may be asked, and the rule might thus be interpreted as preventing the officer not only from asking any further questions, but from invoking any other sort of power to question the voter's competence.⁵³ If, however, registration and presiding officers continue to work on the basis that the common law incapacity rule survives as a supplement to the new legislative framework, some of the potential benefit of the reforms may be lost. Creating a presumption of entitlement to registration which is liable to be frustrated by the application of archaic common law provisions of uncertain extent falls far short of the full enfranchisement of those with psychiatric illnesses, if indeed it is safe to assume that this was the policy aim behind the 2000 Act's provisions.

If it emerges that the common law rule has survived the enactment of the RPA 2000, it might well be susceptible to a successful Human Rights Act challenge. This would again be on the basis of Protocol 1-3, arguing that the electoral rights of a voter with a mental illness are violated by the common law rule, which has the potential to interfere with the right to register and the right to vote. Additionally, it might be argued that the rule clashes with Article 14, the non-discrimination provision, by disfavouring those who share the status of being mentally ill.⁵⁴ This argument could be strengthened by referring to the criterion of competence, demonstrating that the common law rule amounts to a selective application of that criterion to one group of potential voters, while other voters are insulated from its effects. It is highly iniquitous for the law to allow the possibility of a mentally ill person being deprived of the right to vote, while permitting others, whose capacity to exercise political choice has been neither questioned nor tested, to enjoy the right unchallenged. Those whose illness causes their reason to be diminished may display this in easily observable ways, but to take that as a shorthand for their incapacity to cast a vote while assuming the full reasoning capacity of those whose behaviour is outwardly unremarkable is both unreliable as a means of testing competence and an unjustifiable violation of the political rights of persons with mental illness. Nor is it possible to defend the law on the basis that it is necessary to retain some means of

52 See Little, *Schofield's Election Law* n 50 above para 9.14, Home Office circular RPA 379 27 August 1993, 'Electoral Registration of Mentally Ill or Learning Disabled (Mentally Handicapped) People', para 13, directing election officials to use the statutory questions to test the capacity to vote of those suspected of having a 'profound mental disorder'.

53 This would seem to be a reasonable interpretation of rule 35(4): 'Save as by this rule authorised, no inquiry shall be permitted as to the right of any person to vote'.

54 This argument need not depend on establishing that mental illness is a 'status' under Art 14. Even if a court were reluctant to interpret the provision in that manner, it is clear from the text of Art 14 that it provides a list of illustrative examples, rather than an exhaustive list of prohibited grounds of discrimination. The fact that mental illness is not mentioned should not therefore prove a barrier to such a discrimination claim. See *Engel & Others v The Netherlands*, 8 June 1976, Series A, no 22.

restricting the electoral rights of such people in order to ensure that they do not vote at times when their thinking may be disordered by their illness. The weakness in this reasoning is that it overlooks the matter of the law's general unconcern about the level of competence in voting present in the population of those unaffected by mental illness: no competence tests apply to those who may, for reasons of lack of interest or general confusion, have difficulty in understanding the process of voting. The selective competence testing of those with mental illness is at best a random and unprincipled exercise, and at worst an expression of prejudice.⁵⁵ The law ought to be reformed to preclude the possibility entirely.

Encouraging Voting

A principal aim of the RPA 2000 is to increase voter turnout. To that end the Act makes provision for new ways of voting, which it is hoped will encourage electors to cast their ballots.

Pilot Schemes

Section 10 of the RPA 2000 authorises local authorities in England and Wales to submit proposals to the Secretary of State for pilot voting schemes at local government elections.⁵⁶ These schemes may run throughout the entire local authority area, or be restricted to parts of it.⁵⁷ (Scottish local authorities are excluded from this provision because Scottish local election procedure is a devolved matter).⁵⁸ If approved, the Secretary of State must make provision for the implementation of the scheme by order.⁵⁹ The proposed schemes may relate to when and where voting is to take place at local government elections; how the votes cast are to be counted; and the sending by candidates of election communications free of postal charges.⁶⁰ This broad language provides considerable scope for innovation and is supplemented by a subsection stating

55 A related argument which is sometimes raised with respect to voting by prisoners and psychiatric in-patients is that their votes will form a bloc, thus distorting the results of the election. This contention is based on the unlikely presumption that prisoners or patients share a particular political view derived from their experience as prisoners or patients. Even were this shown to be so, (which seems highly unlikely), it is clearly unreasonable and undemocratic to fence such groups out of the political process for fear of the way in which they may vote. The Supreme Court of the United States was presented with just this question in *Carrington v Rash*, 380 US 89 (1965) in which it held that a provision of the constitution of the state of Texas which deprived members of the armed forces settled in the state of the right to vote violated their constitutional rights to equal protection. The state had argued that the collective electoral voice of the soldiers might overwhelm that of the small civilian community near the military base, but the Court's reaction was unequivocal: a state may not fence out a particular section of the electorate because of anxiety about its voting intentions.

56 In addition, the new Electoral Commission will be entitled to participate with a local authority in the preparation of a joint application: Political Parties, Elections and Referendums Bill 2000, cl 9.

57 RPA 2000, s10(4).

58 Scotland Act 1998, Sched 5. (The franchise laws governing local government elections in Scotland are, however, a reserved matter.)

59 RPA 2000, s10(1)(b). The Political Parties, Elections and Referendums Bill 2000 adds s10(1A), which provides that the Secretary of State must consult the Electoral Commission before making an order under s10, unless the proposal is made jointly by the Commission and the authority. A copy of the order is to be sent to the local authority and to the Electoral Commission: RPA 2000, s10(5), as amended by the Political Parties, Elections and Referendums Bill 2000, Sched 21 para 16.

60 RPA 2000, s10(2); s10(3) authorises the Secretary of State to disapply RPA 1983, s75 (prohibition on unauthorised third party election expenditure) in order to permit councils lawfully to fund the distribution of election literature.

explicitly that such schemes may provide for voting to take place on more than one day, and at places other than polling stations.⁶¹ Pilot schemes may not, however, experiment with different voting systems.⁶²

A number of local authorities instituted pilot schemes for the local government elections held in England and Wales on 4 May 2000. Schemes included extending the poll over one week, voting in public libraries and art galleries, 'mobile' polling (eg taking the ballot box to residential homes for the elderly), weekend voting, and voting in shops and supermarkets.⁶³ Experiments with electronic voting and vote counting also took place. The preliminary results did not appear particularly encouraging, with low turnout persisting despite the innovations.⁶⁴ It seems that, although quite popular, the alternative voting methods appealed mainly to voters who would have turned out anyway. One pilot which did, however, seem to be very successful involved the use of all-postal ballots: this experiment saw turnout rise by up to 32% in those wards in which it was tried.⁶⁵

Within three months of conducting such a pilot scheme, the Electoral Commission is required to submit a report on the exercise, prepared in consultation with the local authority, to the Secretary of State.⁶⁶ This is to contain an assessment of the scheme's success 'in facilitating voting at the elections in question ... or in encouraging voting ... or enabling voters to make informed decisions at those elections.'⁶⁷ The Commission must include with its assessment of the first factor a statement of its opinion as to whether voter turnout was higher with the new scheme; whether voters found the procedures easy to use; whether there was any increase in personation or other malpractice; and whether the procedures led to any costs or savings for the authority.⁶⁸ Section 11 provides that the Secretary of State may, in light of such a report, decide to extend the scheme to local government elections throughout England and Wales and that this may be done by a statutory instrument requiring an affirmative resolution of both Houses of Parliament.⁶⁹ This provision attracted particular criticism during the passage of the Bill, concerns centring on the power granted to the Secretary of State to amend primary legislation (the RPA 1983) by order so as to 'roll out' pilot schemes to other

61 RPA 2000, s10(3). Where polling is extended over more than one day, there is a prohibition on the publication of exit polls before the final close of the poll: RPA 2000, Sched 6 para 6 inserting new s66A in RPA 1983. This Lords' amendment was prompted by concerns that interim publication of exit polls might influence voters yet to cast their ballots.

62 A Lords' amendment to permit pilots including systems of voting was not successful, the government taking the view that that matter was best left to primary legislation. See HL Deb vol 609 cols 1081–1083 15 February 2000.

63 See *The Guardian* 25 April 2000 'Armchair rebels threaten Labour hopes'.

64 A total of 32 authorities ran pilot projects. For details of the schemes and their impact on turnout, see House of Commons Library Research Paper 00/53, *The Local Elections and Elections for a London Mayor*, 23 May 2000. Local authorities may be reluctant to run pilot schemes in May 2001 in case the general election is called to coincide with the local elections, thus requiring the cancellation of the pilot schemes, which would be valid only for the local, but not for the general election.

65 See *The Guardian* 15 May 2000, 'Making it easy'. Interestingly, this article reports voters (validly) handing in postal ballots in person at polling stations, and stating that they would have preferred to vote the old way.

66 RPA 2000, s10(6)–(10) as amended by the Political Parties, Elections and Referendums Bill 2000, cl 9 & Sched 21 para 16. The Electoral Commission is established by cl 1 of that Bill. This report is to be copied to the local authority and published within the authority area in such a manner as it thinks fit: s10(10)(b).

67 RPA 2000, s10(7)(c).

68 RPA 2000, s10(8).

69 In addition to a draft of the statutory instrument containing the order, the Secretary of State must lay before each House a copy of the Electoral Commission report(s) submitted under s10: RPA 2000, s11(4).

elections.⁷⁰ The government's defence of the mechanism amounted to a plea that pressures of parliamentary time would make the process of change too slow if primary legislation were needed for every extension of a pilot scheme. In response to these concerns, the government gave an undertaking that the rolling out of pilot schemes would happen only on the recommendation of the Electoral Commission, and an amendment to that effect was included in the Political Parties, Elections and Referendums Bill.⁷¹ In fact, the power granted to the Home Secretary by this provision is significantly more restricted than it might have been: as originally drafted, the Bill permitted pilot schemes to be rolled out by delegated legislation beyond local government elections to Parliamentary, European Parliamentary, Scottish Parliamentary, Welsh Assembly and Northern Irish Assembly elections. Widespread unease within Parliament about the scope of this proposed power led to a government amendment in the Lords limiting it to local elections throughout England and Wales.⁷² Primary legislation will therefore be required to apply successful schemes to parliamentary and other national elections.⁷³

Absent Voting

The RPA 1983 (and the regulations made thereunder) and the RPA 1985 contained complex rules governing entitlement to an absent ballot, which arguably constituted a disincentive to those considering applying for a postal or a proxy vote. This has been addressed by section 12 of and Schedule 4 to the RPA 2000, which introduce a new scheme whereby any voter may seek a postal vote simply by applying to the electoral registration officer. These provisions apply to parliamentary elections (except in Northern Ireland) and to local government elections throughout the UK.⁷⁴ The application may request a postal ballot for a particular elections, for all elections in a given period, or indefinitely. The registration officer must grant the application if satisfied that the applicant is a registered voter and that any prescribed requirements concerning the form of the application (eg a deadline) have been met.⁷⁵ This relaxation of the rules will permit voters to request postal ballots although they will not necessarily be away from home on polling day. Proxy votes, as before, will only be available to those voters who cannot reasonably be expected to go to the polling station in person.⁷⁶

These reforms appear to be predicated on an assumption that low voter turnout is attributable to individuals finding the costs of voting too high.⁷⁷ There is, however,

70 See eg HL Deb vol 609 cols 1088–1097 15 February 2000.

71 Political Parties, Elections and Referendums Bill 2000, cl 9 & Sched 21 para 17, adding to s11 the condition that 'The power of the Secretary of State to make such an order shall, however, be exercisable only on a recommendation of the Electoral Commission.'

72 HL Deb vol 610 cols 829–830 6 March 2000. This change had also been recommended by the Lords Delegated Powers and Deregulation Committee, 4th Report, 9 February 2000, paras 18–27.

73 There have been press reports that the government is considering such a move with respect to postal voting, see *The Independent* 1 June 2000 'Postal votes for all may be allowed at next general election, says Straw'.

74 The provisions apply only to local elections in Northern Ireland because of concerns about the potential for their abuse during parliamentary elections. RPA 2000, s12(2)(b). The administration of Scottish local elections (a devolved matter under the Scotland Act 1998) is covered in response to a decision of the Scottish Parliament to that effect.

75 RPA 2000, Sched 4, para 3(1). Regulations will prescribe the form of the application.

76 RPA 2000, Sched 4, para 3(3). Only minor changes have been made to the rules governing proxy voting due to concerns that any relaxation of the rules would create a heightened risk of fraud.

77 The 'costs' of voting relate not only to the time and effort involved in the act of casting a ballot, but also to the process of informing oneself about the issues and candidates prior to the poll.

little evidence available to demonstrate that the fairly minimal costs of casting a vote in the present system *are* generally regarded by non-participants as being too high. Even if such evidence were available, we would need to know the precise nature of the costs deterring non-participants. Those costs might be related to the physical process of travelling to and from a polling station, in the rain, during the half hour usually reserved for watching a TV show. Extending polling hours (or days) and offering more postal votes might then improve participation rates. But such measures will do little to address the deeper problems underlying low participation. It would seem reasonable to suppose that many non-participants abstain in consequence of their sense of alienation from the political process which they take the ballot to represent. A further number might characterise non-participation as a positive act, a silent expression of their desire to be dissociated from the political sphere. Others might regard non-voting as a privilege they are free to enjoy for as long as the governors and legislators whom the participants pick continue to rule in a manner which they find satisfactory or, at least tolerable. The pilot schemes and other measures contained in the Act are unlikely to reach these groups of non-participants.

It is notable also that hardly any consideration was given during the pre-legislative and legislative processes to the possibility of instituting some form of compulsory voting in response to declining rates of participation.⁷⁸ The traditional objection to compulsion is that it strips individuals of their right not to vote: voting, which lies at the heart of democratic structures created to preserve freedom, should not itself be a forced act. Philosophical answers to such objections are possible, however, and laws compelling voting do operate successfully in several developed democracies around the world.⁷⁹ If participation rates continue to fall, it may become increasingly difficult for governors and legislators to defend their unreasoned rejection of the compulsory voting option.

Conclusion

The RPA 2000 is a response to some of the most obvious defects and limitations of current electoral law and it provides useful, if somewhat patchy, solutions to those problems. The system of electoral registration has been modernised, the electorate subtly re-defined and new schemes of voting offered. The Act's underlying weakness is that it is excessively orientated towards increasing electoral participation rather than towards the articulation of principles governing the extent of rights to, and the conditions of, electoral participation. The reforms in the RPA 2000 happen to offer a (slightly) wider franchise and varied methods of voting because those measures are perceived to be means to its particular goal. In an altered political climate, however, in which low voter turnout was not regarded

78 Although the Home Affairs Select Committee did not favour compulsion as an option, it did express the view that there should be a public debate on the matter: Fourth Report n 3 above paras 26–29. There are now indications that some members of the Government are pressing for serious consideration to be given to the option, see *The Scotsman* 29 May 2000 'Labour to consider mandatory voting as way to reverse decline in turnout'; *The Independent* 23 May 2000, 'Postal ballots plan to halt drop in turn-out'.

79 eg Australia, Italy, Belgium, Luxembourg and Greece. See A. Lijphart, 'Unequal Participation: Democracy's Unresolved Dilemma' (1997) 91 Am Pol Sci Rev 1; M. Feeley, 'A Solution to the "Voting Dilemma" in Modern Democratic Theory', (1974) 84 *Ethics* 235; R.L. Hasen, 'Voting Without Law', (1996) 144 U Pa L R 2135.

as a particular political concern, laws restricting the franchise might as easily be enacted in pursuit of some other aim (such as, for example, discouraging immigration by withdrawing the right to vote of all but UK citizens). This remains a possibility in part because the law, including this Act, has so far failed to offer a complete and coherent account of the criteria governing enfranchisement and their interrelationship.

The RPA 2000 has stumbled onto some positive-sounding answers, but it is not at all clear that the correct questions were posed in the first place. Decisions about who is entitled to vote and how the right to vote is exercised ought to involve a systematic consideration of the basic principles relevant to enfranchisement and participation. Do we, for example, wish to retain citizenship as the basic criterion of enfranchisement, and if so, what forms of citizenship should qualify for full suffrage? Commonwealth? British? European Union? To what extent is the notion of competence to vote a relevant one, and in what sorts of laws might it find legitimate expression? If the notion of being a good citizen is relevant to the distribution of voting rights in the 21st century, as appears to be evidenced by the continuing disenfranchisement of imprisoned convicts, then why should that rule not apply also to those convicted of more minor crimes and to those who escape imprisonment? Does serial non-voting represent a lack of the sort of virtue on which the system appears to condition retention of the right to vote? Might the government legitimately respond to mass non-participation by invoking the law to compel voting? These and other questions are left unanswered by the 2000 Act. A fundamental reconsideration of the principles on which the law governing the franchise is based is overdue. Equally, the causes and possible responses to mass non-participation at elections need to be more fully explored before effective remedial measures can be implemented. The RPA 2000 offers encouraging glimpses of a radically reformed electoral law but ultimately disappoints by failing to develop more fully the democratic commitments to which it cautiously refers. Like its predecessors, this Representation of the People Act is a partial and pragmatic revision of electoral law, tethered only loosely to principle and tendering at best a sort of democratic development by default.

REPORTS

The Royal Commission on Reform of the House of Lords: A House for the Future?

Meg Russell and Richard Cornes***

The Royal Commission on the Reform of the House of Lords, chaired by Lord Wakeham, published its report in January 2000.¹ The Commission's purpose was to consider the long term reform of the United Kingdom (UK)'s upper house. Its establishment was announced in December 1998, when the government published its House of Lords Bill to remove the hereditary peers from the chamber, and a White Paper setting out its longer term intentions. The terms of reference for the Royal Commission, set out in the White Paper, were as follows:

Having regard to the need to maintain the position of the House of Commons as the pre-eminent chamber of Parliament and taking particular account of the present nature of the constitutional settlement, including the newly devolved institutions, the impact of the Human Rights Act and developing relations with the European Union:

- to consider and make recommendations on the role and functions of a second chamber;
- to make recommendations on the method or combination of methods of composition required to constitute a second chamber fit for that role and those functions;
- to report by 31 December 1999.²

These were ambitious targets. The Commission did not meet until March 1999 so had only nine months to report. It was required to take into account the raft of constitutional changes put into place since 1997, but to a large extent the outcomes of these constitutional changes were impossible to predict. The 'newly devolved institutions' in Scotland and Wales were elected two months after the Commission first met, and only assumed their powers two months after that. The Human Rights Act came into force across the UK only in October 2000, ten months after the Commission had reported. The timescale allowed little time for either public consultation or background research. In particular the Commission had little opportunity to draw on the experiences of overseas constitutions, many of which already exhibit characteristics which the UK is likely to develop as a result of these reforms. It seems impossible that any commission, in these circumstances, could 'take account of the present nature of the constitutional settlement', as the Royal Commission was required to do. Indeed an investigation into the role, functions and composition of the second chamber requires some fundamental constitutional questions to be addressed, amongst them the relationship between parliament, the executive and the judiciary, the relationship between the centre and the territorial administrations in a devolved

* The Constitution Unit, School of Public Policy, University College London.

** School of Law, University of Essex.

1 *A House for the Future*, Royal Commission on the Reform of the House of Lords Cm 4534 (2000).

2 *Modernising Parliament – Reforming the House of Lords*, Cm 4183 (1999).

© The Modern Law Review Limited 2001 (MLR 64 1, January) Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF and 350 Main Street, Malden, MA 02148, USA

Britain and even the relationship between church and state. The Royal Commission had neither the resources, the time, nor the political backing to carry out such an investigation.

This paper analyses the Royal Commission's recommendations, with particular reference to the new constitutional settlement. Our analysis is informed by lessons which can be drawn from other comparable democracies. We conclude that the Commission's proposals were timid, and did not adequately take into account the ways in which the UK constitution is liable to develop. This is unsurprising given the constraints within which it operated. The Commission also failed to tackle adequately the question of the appropriate powers for a second chamber in modern Britain, and the way in which the composition of the chamber will govern its ability to effectively use its powers. Given the removal of the hereditary peers from the chamber in 1999, the context for debate about its reform has already radically changed. The issues outlined in this paper are some of those which are likely to come to the fore as this debate develops.³

The powers of the chamber

The Royal Commission was not explicitly required to consider the powers of the reformed upper house, only its 'role and functions'. However, inevitably it was impossible to separate these issues (nor the issue of composition), and the Commission made a number of recommendations on powers. In doing so, however, the Commissioners clearly felt constrained by the requirement to 'maintain the position of the House of Commons as the pre-eminent chamber'. Here we consider the Commission's recommendations with respect to four different areas.

Making and breaking governments

The House of Lords currently has no direct part in the process that leads to the formation of government following an election. There is also no concept of a confidence vote in the upper house. The British parliamentary system gives sole power to the House of Commons in the making and breaking of governments. This is entirely appropriate in a system where the House of Commons is elected and the House of Lords is not. However, the lack of power over government formation is one of the defining features of upper chambers of parliament worldwide. Only in Italy does government face confidence motions in both the upper and lower houses. The notion of giving a reformed UK upper house the right to remove government from office was not even considered in the Royal Commission's report. It is, however, introduced here for an illustrative purpose. For a narrow interpretation of the 'pre-eminence' of the House of Commons could rest on this characteristic alone. The Royal Commission chose to interpret the pre-eminence condition far more broadly, with consequences for their recommendations in each of the three areas discussed below.

³ We do not seek to provide an exhaustive critique of all the Commission's proposals. In particular we do not address the issues of religious representation in the chamber, links to the peerage, the name of the new chamber, or changeover arrangements. Discussion of all of these issues may be found in the Commission's report.

Ordinary legislation

The powers of the House of Lords over ordinary legislation are set out in the Parliament Acts 1911 and 1949. In summary the chamber can delay most ordinary legislation for approximately one year, and 'Money Bills' for one month.⁴ The category of Money Bills is defined narrowly as only those dealing exclusively with 'charges'.⁵ Money Bills are identified by the Speaker of the House of Commons, whose word is final. The Royal Commission proposed no change to these powers. This would leave the UK upper house with moderate delaying powers in international terms. This is illustrated in Table 1.

Table 1: Power of the upper house over ordinary and financial Bills in 10 democracies

Country	Power over ordinary Bills	Power over financial Bills
Australia	Total veto	As for ordinary Bills
Canada	Total veto	As for ordinary Bills
France	6–12 months' delay	As little as 15 days' delay
Germany	Total veto over Bills affecting states, delay over others	As for ordinary Bills
Ireland	Three months' delay	Three weeks' delay
Italy	Total veto	As for ordinary Bills
Japan	Two months' delay	One month's delay
Spain	Two months' delay	As for ordinary Bills
Switzerland	Total veto	As for ordinary Bills
UK	Approximately one year's delay	One month's delay for 'Money Bills'
USA	Total veto	As for ordinary Bills

The Bryce Commission, established in 1917 by David Lloyd George to study options for House of Lords reform, wisely stated that one of the four main functions of an upper house was 'The interposition of so much delay (and no more) in the passing of a Bill into law as may be needed to enable the opinion of the nation to be adequately expressed upon it.'⁶ The chamber's current powers could be said to meet this requirement. A veto power obviously puts an upper house in a stronger position, but can easily be seen as encroaching too much on the powers of government and the lower house. A delay power of only two or three months, such as that enjoyed by some upper houses, may on the other hand be insufficient to force government to think again or to allow public debate.

Whilst its formal powers are set down in statute, the House of Lords' powers are restricted in practice by many conventions. The best known is the so-called 'Salisbury convention', agreed by Conservative leader in the Lords Lord Salisbury following the 1945 Labour landslide. This convention acknowledged the rights of the elected government by stating that Bills implementing manifesto commitments would not be rejected outright by the Conservative-dominated House of Lords. The convention endured throughout the remainder of the century, through periods of both Conservative and Labour government, and thus acknowledged not only the

4 s 2(1) of the 1949 Act.

5 s 1(2) of the 1911 Act.

6 *The Reform of the Second Chamber* Cd 9038 (1918) para 6.

party differences between the chambers but also the greater legitimacy of the elected house.⁷ In fact the powers of the chamber have been used only rarely in the last century, due to the legitimacy problems suffered by the hereditary and Conservative-dominated upper house. It is thus impossible to separate the issues of composition, legitimacy and powers of the chamber. This matter is discussed in greater detail towards the end of the paper.⁸

Whilst proposing that the upper house's powers over legislation remain unaltered, the Royal Commission proposed that a joint committee of the two chambers might be established, in order to negotiate amendments in the event of any disputes between them. With a more activist upper house such a committee could be a useful addition to procedure. Joint mediation committees are common overseas, and the introduction of a committee in the UK was suggested by the Bryce Commission in 1918. The Royal Commission does not specify the detail of how such a committee should be organised, save for suggesting that its members should be senior and reflect the party balance of each chamber. In practice the success of such committees also depends on a number of other factors. For example the committee in the United States gives equal weight to representatives of both chambers by requiring a majority of each group to support any proposals it makes. The committee in Germany benefits by meeting in private, and having freedom to trade amendments on any part of the Bill it is considering. The committee in France, in contrast, is ineffective because it tends to split along party lines, and in any case can be overruled by government or the lower house.⁹ The Royal Commission's open-ended suggestion would leave the detailed design of a conciliation committee to parliament itself, and this is one of the issues which should be discussed by the joint committee established to look at the 'parliamentary implications' of the Commission's proposals.¹⁰

Constitutional matters

The Royal Commission states that 'One of the most important functions of the reformed second chamber should be to act as a "constitutional long-stop"' (Recommendation 15). This reflects the proposal of the Bryce Commission that the power of delay 'would be specially needed as regards Bills which affect the principles of the Constitution'¹¹ and the established role played by many second chambers around the world. The House of Lords has to some limited degree attempted to act as a constitutional guardian.¹² It also has a limited constitutional power, to veto a Bill extending the life of a parliament. The Commission, however, did not take the opportunity to propose that the upper house be given significant

7 The Royal Commission proposes that the Salisbury convention be retained in a reformed upper house, although they reject putting it on a more formal basis (paras 4.21–4.24). In practice, however, it is difficult to see how such a convention will continue to be respected in an upper chamber which has more confidence and legitimacy.

8 For a fuller discussion of the interplay between these three issues see M. Russell, *Reforming the House of Lords: Lessons from Overseas* (Oxford: Oxford University Press, 2000) ch 12, and A. Lijphart, *Democracies: Patterns of Majoritarian and Consensus Government in Twenty-one Countries* (New Haven: Yale University Press, 1984) ch 6.

9 For a fuller discussion of the role of joint mediation committees see M. Russell *ibid* chs 6 and 13.

10 For the announcement of this committee see HL Deb vol 614 WA col 122 3 July 2000.

11 n 6 above para 6.

12 For example, for an account of the chamber's intervention in the abolition of the Greater London Council in 1986 see D. Shell, *The House of Lords* (London: Harvester Wheatsheaf, 1992) 168–170. For its support of legislative initiatives proposing a Bill of Rights see A. Reidy, *The House of Lords: In Defence of Human Rights* (London: Constitution Unit, 1999).

new constitutional powers. Instead the chamber would exercise its proposed new constitutional watchdog role through a new set of scrutiny committees (described on pages 91–92 below).

The UK is one of only three Western democracies without a written constitution.¹³ This makes it more difficult to give the upper chamber special powers over constitutional legislation than it is in other bicameral states. Nonetheless, the raft of constitutional changes since 1997 mean that it would be beneficial to give the upper house an enhanced role in constitutional protection. This could be used to ensure that there is some stability for the newly developing constitution, and that any future changes should have broad support. Table 2 shows the procedure for constitutional change in the same 11 countries which were illustrated in Table 1. This shows that in all 10 overseas states there is a strict procedure for constitutional change which is more demanding than that for ordinary legislation. Unless a referendum is required, approval by the upper house is essential – often by a qualified majority. Notably, even those upper chambers which have relatively weak powers with relation to ordinary legislation, such as those in France, Spain and Japan, have an absolute veto over proposed constitutional amendments.

Table 2: Procedures for constitutional change in 10 democracies

Australia	Must be passed by referendum.
Canada	Must be passed by provincial assemblies.
France	Must pass both chambers, then either joint sitting by 3/5 majority, or a referendum.
Germany	Must pass both chambers by 2/3 majority.
Ireland	Must be passed by referendum.
Italy	Must pass both chambers by absolute majority, and if not by 2/3 majority referendum may be called by 1/5 of members of either house, 500,000 electors, or five regional assemblies.
Japan	Must pass both chambers by 2/3 majority.
Spain	Must pass both chambers by 3/5 majority, or lower house by 2/3 majority and upper house by absolute majority. Referendum may be requested by 1/10 of members of either house.
Switzerland	Must pass both chambers, else referendum.
UK	Treated as ordinary legislation.
USA	Must pass both chambers by 2/3 majority, plus states' approval by referendum.

The Labour Party's 1992 general election manifesto committed it to creating a second chamber which could 'delay, for the lifetime of a parliament, changes to designated legislation reducing individual or constitutional rights'. This commitment had been dropped by 1997. The Liberal Democrats supported a similar arrangement in their evidence to the Royal Commission, stating that 'the reformed second chamber, as guardian of the constitution, should have the right to reject any bills which are designed to make significant changes to the constitution'. Similar proposals have also been made by British constitutional lawyers.¹⁴

¹³ The other two are Israel and New Zealand, both of which have single chamber parliaments.

¹⁴ See, for example, R. Blackburn, 'The House of Lords' in R. Blackburn and R. Plant (eds), *Constitutional Reform* (London: Longman, 1999).

The approach supported by the Labour Party in 1992 would give the upper house greater power over amendment of a certain number of existing Acts. These might include, for example, the Scotland Act, Government of Wales Act and Human Rights Act. Requiring a special procedure for amendment of these Acts would acknowledge their status as part of the UK's increasingly codified constitution. In practice this could be achieved through naming the Acts in the Bill which introduced the second stage of House of Lords reform. The chamber's veto over a Bill to extend the life of a parliament is entrenched in this way by the two Parliament Acts.¹⁵ The Royal Commission accepted the principle of extending the upper house's veto, by suggesting that this should apply to any future Bill seeking to amend its own powers (Recommendation 19). However, this arrangement would not prevent the outright abolition of the upper house by the House of Commons alone. Neither would it prevent the abolition of other institutions such as the Scottish Parliament and Welsh Assembly, or fundamental alterations to their powers, by the lower house at Westminster.

If the new upper house were to be given special powers over a designated number of constitutional Acts, these could take many forms.¹⁶ The chamber might be given either a total veto over such legislation, or an extended power of delay (in 1992 the Labour Party favoured the latter). Such a mechanism would be strengthened by a requirement for a qualified majority in the upper house, as applies in many overseas parliaments. An alternative, perhaps more in line with recent British tradition, would be to adopt a system similar to that in Switzerland whereby government would need to seek approval in a referendum for any constitutional reform which did not receive the consent of the upper house.¹⁷

This approach, however, would give rise to a number of practical complications, which the Commission points out. One difficulty is that no list would be complete, and it might be necessary to make many future amendments in order to give protected status to new 'constitutional' Acts. More serious would be the problem that relatively minor measures will require consequential amendments to, for example, the Scotland Act. If the upper house enjoyed a veto power over any such amendments this could give it far-reaching and unintended powers.

An alternative approach to this peculiarly British problem would be to establish a procedure to distinguish any future Bills as constitutional. This was the approach suggested by the Liberal Democrats in 1999. The upper house could then have increased powers over any Bill designated in this way. The precedent would be the current designation of 'Money Bills' by the Speaker of the House of Commons, which results in reduced power for the Lords over such Bills. The designation of constitutional Bills might likewise rest with the Speaker, within a broad definition which might be included in a new House of Lords Act.¹⁸ However, it seems more

15 See n 4 above.

16 For a discussion of whether parliament can create such restrictions for its successors, see S. de Smith and R. Brazier, *Constitutional and Administrative Law* (London: Penguin Books, 1998). De Smith and Brazier conclude that 'If Parliament can make it easier to legislate, as by passing the Parliament Acts or abolishing the House of Lords, it can also make it harder to legislate' (96).

17 The Liberal Democrats proposed to the Royal Commission that this was the process which should apply to those Bills designated as constitutional.

18 Such a definition might include, for example, legislation that alters: the status of the monarchy, the structure or powers of the legislature, the executive, and the judiciary, the existence and power of devolved assemblies, and the basic rights as set out in the Human Rights Act 1998. A broader definition might also include Bills which alter: the control and organisation of political parties, the voting system, conduct of elections and referendums, the administration of justice, and the administration of local government.

likely that there would be disputes over the classification of such Bills than over Money Bills, thus an appeal mechanism might be necessary. In overseas systems the final word on such matters would be given to the Constitutional Court.¹⁹

The practical obstacles were used by the Royal Commission as partial justification for not extending the chamber's role over constitutional matters. However, one of the Commission's main justifications was that 'increasing the powers of the second chamber over any particular category of legislation would be inconsistent with maintaining the position of the House of Commons as the pre-eminent chamber of Parliament' (para 5.7). This is one example of the Commission interpreting 'pre-eminence' too broadly, in our view – particularly if the chamber's powers over government and over ordinary Bills are limited in the ways outlined above.

Delegated legislation

The Parliament Acts do not limit the House of Lords' powers over delegated legislation, which was then not such an important feature of parliamentary work as it is now.²⁰ Consequently the chamber retains a total veto over all such legislation. One of the conventions of the House of Lords has been that this veto is not used. The chamber has generally followed a self-denying ordinance and stopped short of voting down orders which it believes to be problematic. Its influence has generally been applied less formally, through government taking account of objections raised in Lords debates, or even through motions calling for Instruments to be withdrawn.²¹ During the twentieth century there was just one breach of the convention, when the chamber voted down the Southern Rhodesia Sanctions Order in 1968.²²

However, recent reform of the chamber, and its greater sense of its legitimacy, has resulted in this convention being brought into question. In a speech in November 1999 Lord Strathclyde, Conservative leader in the Lords, stated that '[m]ost of the conditions that gave rise to Salisbury have gone', and that as far as the convention on secondary legislation is concerned 'I declare this convention dead'.²³ Indeed on 22 February 2000 the chamber used its power to veto the Greater London Authority (Election Expenses) Order, which forced government to rethink the arrangements for availability of free postage of campaign literature to London Mayoral candidates.

In recognition of the fact that a more legitimate upper house might seek to use its powers more frequently, the Royal Commission proposed that the chamber's power over delegated legislation be reduced from total veto to a brief power of delay. Following a rejection by the upper house the House of Commons would be able to reaffirm its support for an Instrument at any time within three months, thus

19 For example in Germany the Constitutional Court adjudicates on disputes over the status of Bills, which occur because the upper house has a veto over only Bills affecting state finances or responsibilities. Frequent referrals are made to the Court by the upper house. See C. Jeffery, 'German Federalism in the 1990s: On the Road to a "Divided Polity"?' in K. Larres (ed), *Germany Since Unification* (Basingstoke: Macmillan, 1998).

20 In 1947–49 when the new Parliament Bill was under discussion, consideration was given to including new provisions for delegated legislation, but this idea was rejected. See J. Griffith and M. Ryle, *Parliament: Functions, Practice and Procedures* (London: Sweet & Maxwell, 1989) 488.

21 D. Shell, *The House of Lords* (London: Harvester Wheatsheaf, 1992) 215–223.

22 See J. Morgan, *The House of Lords and the Labour Government 1964–1970* (Oxford: Oxford University Press, 1975) ch 5.

23 Speech hosted by Politeia, 30 December 1999

overriding the chamber's concerns. Under such arrangements the reformed upper house might intervene more frequently, but would have little real power. There would thus be a real danger that the chamber would not hold sufficient threat to require government to rethink.

Particular concerns have been raised about the impact which the chamber's reduced power in this area could have over the agreement of remedial orders under section 10 of the Human Rights Act.²⁴ If the courts make a 'statement of incompatibility' on the basis that an Act breaches European Convention rights, a minister may issue such an order to rectify the incompatibility where there are 'compelling reasons' to do so. This action may include repeal or amendment of previous legislation. Remedial orders have the status of delegated legislation, thus under current arrangements the upper house has a power of veto. However, the Royal Commission's proposals would reduce this to a delay which could in practice amount to only a few days or hours. The Lords' veto over remedial orders was part of the accepted framework during parliamentary debates on the Human Rights Act. The Royal Commission considered the arguments for retaining greater upper house powers in this area (para 7.39) but there are concerns that these were dismissed too lightly.

The parliamentary functions of the chamber

The Royal Commission assumed that many of the chamber's parliamentary functions would remain largely unchanged. These include many functions for which the House of Lords is currently well respected, including detailed legislative scrutiny, scrutiny of draft European legislation, and work of specialist committees. At least until such time as the House of Commons takes greater responsibility in these areas, such work by the upper house will remain essential. In this section we concentrate on just three areas where the Royal Commission either proposes that the parliamentary functions of the chamber are changed significantly, or could have done so. These are its relationship with government, its possible territorial role, and other constitutional roles.

Relationship to government

Currently around 12 per cent of government ministers are drawn from the House of Lords. Generally just two cabinet ministers (Leader of the House and Lord Chancellor) are members of the upper house.²⁵ Lords ministers are required to cover all government business in the House, including questions, passage of Bills, statements, and replying to debates.

The Royal Commission proposes that this relationship remains largely unchanged, although it is suggested that standing orders might be amended to allow ministers who are members of the House of Commons to make statements and answer questions in the upper house. This would have the advantage that members of the chamber could be addressed by the minister with prime responsibility for a policy area, or overall control of a government department,

24 See R. Blackburn 'How Powerful Should the Second Chamber Be?' in *Conference Papers: The Future of the House of Lords*, (London: Constitution Unit, 2000)

25 Since the 1950s Conservative cabinets have never included more than four peers, and Labour cabinets not more than two (Shell, n 12 above, 74–75).

rather than facing, as frequently is the case, a Lords minister who is required to cover too wide a range of briefs. This could potentially improve the accountability of government to the upper house. However, if Commons ministers are to gain access to debates in the upper house, this raises the question of why it is necessary to continue to appoint Lords ministers. Currently the limitation on such ministers comes partly as a result of the perceived legitimacy of the unelected and, until recently, hereditary-dominated upper house. In a reformed chamber with greater legitimacy this restriction would be to some extent be lifted.

One of the concerns frequently expressed about inclusion of elected members in the upper house is that the chamber becomes less distinct from the House of Commons. The Royal Commission was, explicitly and correctly, concerned to protect the distinct nature of the chamber and independence of its members, so that 'even those members of the second chamber who are affiliated to a party should be prepared to deal with issues on their merits and should exercise a certain independence of judgement' (para 10.27).

One way of reinforcing the distinction between the two chambers would be to restrict the appointment of ministers to the lower house only. This was the policy of the Labour Party in the run up to the 1992 election²⁶ and was proposed by the Liberal Democrats in their evidence to the Royal Commission. It has also been suggested by others, including Conservative MP Andrew Tyrie.²⁷ Since government can be made or broken by the House of Commons alone, such a restriction would underline the role of the Commons as the 'government chamber'. It would thus reinforce the position of the upper house as the chamber of review, more independent of government and party influence. In particular, the removal of the ministerial career ladder from the upper house would encourage more independent behaviour amongst its members.

In parliamentary systems overseas it is usual for most ministers to be drawn from the lower house. In some cases – such as that of South Africa – the constitution forbids appointment of ministers from amongst upper house members. However, it is also usual for ministers to appear in both chambers, irrespective of in which chamber (if either) they hold a seat.²⁸ Implementing such a change in the UK would therefore be consistent with international practice, whilst potentially improving the parliamentary accountability of government and independence of the upper house.

Territorial functions

The Royal Commission was explicitly required to take into account the state of devolution in the UK when drawing up its recommendations. Devolution weakens the legal unitary nature of the UK and will increase the importance of territorial politics and inter-territorial tensions.²⁹ However, this was a difficult issue for the Commission to consider fully, given the under-developed nature of devolution at the point that they met. One of the classic functions of an upper house is to act as a chamber of territorial interests, to diffuse territorial tensions and provide a forum in

26 *Meet the Challenge, Make the Change* (1989) stated that 'Ministers will not sit in the second chamber'.

27 A. Tyrie, *Reforming the Lords: A Conservative Approach* (London: Conservative Policy Forum, 1998).

28 See Russell, n 8 above, ch 9.

29 R. Hazell, 'The New Constitutional Settlement' in R. Hazell (ed), *Constitutional Futures* (Oxford: Oxford University Press, 1999).

which they may be resolved.³⁰ All federal bicameral states worldwide base representation in the upper house on regions, states or provinces, whilst the same is also true of a quarter of unitary bicameral states.³¹ The Royal Commission concluded that '[t]he reformed second chamber should be so constructed that it could play a valuable role in relation to the nations and regions of the United Kingdom, whatever pattern of devolution and decentralisation may emerge in future' (Recommendation 25). However, whilst the Commission made some limited proposals on territorial representation – described later in this paper – it did little to identify what territorial role or functions the upper house might play in a devolved UK.

The Commission's main proposal in terms of territorial functions is that there should be a Devolution Committee in the upper house, which would consider relations between the devolved institutions and the centre, and relations between the institutions themselves. This committee could be useful, but its narrow remit is not sufficient. A stronger mechanism would be to look beyond the technicalities of devolution itself and build the territorial function into the everyday work of the upper house, considering for example the impact of all Westminster Bills on the different parts of the UK. Such scrutiny might avert future claims that government policy was fuelling a 'north-south divide'. In the German upper house (the Bundesrat) all policy committees have a territorial format, comprising one representative from each of the 16 states. In contrast the Spanish upper house has one territorial committee which scrutinises Bills of particular regional interest.³² In a reformed upper house with a truly territorial nature, either of these models might be adopted. However, their success would depend on the composition of the chamber and the links its members could legitimately claim with the politics of the UK's nations and regions.

The Commission explicitly rejected proposals that members of the upper house should represent devolved institutions rather than citizens. It feared that members chosen in this way would become 'delegates of the bodies that elected them, voting according to instructions rather than conscience' (para 6.17). However, another danger is that members representing the nations and regions – such as those proposed by the Commission (see below) – will become delegates of their parties, with no real link to the institutions in their nations and regions. This is seen in many federal second chambers, such as those in Australia and Canada.³³ A possible solution would be to require members representing the nations and regions to make regular reports to their respective assemblies or parliaments – answering questions in the chamber or accounting to its committees. Such arrangements could start in Scotland, Wales and Northern Ireland, spreading to England if regional assemblies are established. This would build in a degree of accountability, and formal links between the institutions.³⁴

Other constitutional roles

The Royal Commission proposes the establishment of a Constitutional Committee in the upper house, of which the Devolution Committee would be a subcommittee.

30 G. Tsebelis and J. Money, *Bicameralism* (Cambridge: Cambridge University Press, 1997).

31 D. Derbyshire and I. Derbyshire, *Political Systems of the World* (Oxford: Helicon, 1996) 59.

32 See Russell, n 8 above, ch 10; C. F. Juberías, 'A House in Search of a Role' in S. C. Patterson and A. Mughan (eds), *Senates: Bicameralism in the Contemporary World* (Columbus, Ohio: Ohio State University Press, 1999).

33 See Russell, n 8 above, ch 10.

34 See M. Russell, 'The Territorial Role of the Upper House' *Journal of Legislative Studies* (forthcoming, 2001).

A Human Rights Committee would also be established, probably as another sub-committee. These arrangements would formalise the role which the second chamber has long seen for itself as a constitutional guardian, but without the formal powers which would normally accompany such a role (see above). The inclusion of a constitutional committee amongst the bodies in the upper house is relatively common overseas. This would be a welcome addition to the work of the UK upper house, which would complement the House of Commons and could help highlight potential constitutional problems before they occur. As the government adjusts to new constitutional arrangements this could be a valuable role. The House of Lords has since agreed that such a committee will be set up in advance of the other Royal Commission proposals being implemented.³⁵

The Human Rights Committee proposed by the Royal Commission would complement the work of the new Joint Committee on Human Rights, comprising members from both chambers. The Joint Committee will have a limited remit and will automatically scrutinise only those Bills where government has not been able to issue a statement of compatibility with the ECHR.³⁶ Such Bills are likely to be rare. The proposed upper house committee would instead consider the human rights implications of all Bills. Similar committees exist overseas, including a particularly effective pair of committees in the Australian Senate, which scrutinise all Bills and Statutory Instruments against human rights criteria, despite that country's lack of a Bill of Rights.³⁷ A key purpose of such a committee in the UK, as recognised by the Royal Commission, would be to ensure that the human rights implications of legislation are identified before it is passed, rather than afterwards by the courts. This could prove to be an important service once the Human Rights Act is in force. Where problems were identified by the committee, government would be strongly advised to resolve them, even if the upper house did not have the formal power to impose its views.³⁸

Judicial functions

The Law Lords in the upper house

The Royal Commission recommended that 'the Lords of Appeal in Ordinary should continue to be *ex officio* members of the reformed second chamber and carry out its judicial functions' (Recommendation 57). This was based on a conclusion that 'there is no reason why the second chamber should not continue to exercise the judicial functions of the present House of Lords' (para 9.5). Considered from the perspective of what is in the best interests of the Appellate Committee, as a court, this conclusion is too sweeping. While noting the potential application of Article 6 (1) of the European Convention on Human Rights to the

35 HL Deb vol 615 cols 585–589 17 July 2000.

36 In a statement to the House of Commons announcing the establishment of the committee, Leader of the House Margaret Beckett said that the government 'envisage that the Joint Committee's terms of reference will include the conduct of inquiries into general human rights issues in the United Kingdom, the scrutiny of remedial orders, the examination of draft legislation where there is doubt about compatibility with the ECHR, and the issue of whether there is a need for a human rights commission to monitor the operation of the Human Rights Act' HC Deb vol 322 col 604 14 December 1998.

37 See Russell, n 8 above, ch 6.

38 For further discussion see R. Blackburn, 'A Parliamentary Committee on Human Rights' in Blackburn and Plant (eds), n 14 above.

position of the Lord Chancellor, the Commission does not refer to Article 6 (1) in relation to the Law Lords.³⁹

Since the Royal Commission reported, the *McGonnell* case has been decided against the United Kingdom. *McGonnell* is not the only relevant case however. The earlier case of *Procola v Luxembourg* is also pertinent.⁴⁰ These two cases and Article 6(1) are just as relevant, if not even more so, to the Law Lords as they are to the Lord Chancellor. The interpretation of Article 6(1) in both *Procola* and *McGonnell* indicates that while it may not require a strict separation of judicial and legislative powers *per se* – rather that each case must be looked at individually to consider whether there is an Article 6(1) breach – there will be very few cases where a Law Lord who has taken part in legislative debate on a matter can then sit judicially.

Sir John Laws, sitting as an *ad hoc* judge in *McGonnell* said: 'Where there is no question of actual bias, our task under Article 6(1) must be to determine whether the reasonable bystander – a fully informed layman who has no axe to grind – would on objective grounds fear that the Royal Court lacks independence and impartiality'.⁴¹ However, the rest of the Court stated: 'Any direct involvement in the passage of legislation, or of executive rules, is *likely* to be sufficient to cast doubt on the judicial impartiality of a person subsequently called to determine a dispute over whether reasons exist to permit a variation from the wording of the legislation or rules at issue'.⁴² The earlier case of *Procola* sets out a much stricter approach to Article 6(1): 'The mere fact that certain persons successively performed these two types of function [the prelegislative review and hearing an appeal concerning the same legislative instrument] in respect of the same decisions is capable of casting doubt on the institution's structural impartiality'.⁴³ The statement in *Procola* is far closer to requiring a strict separation of powers *per se* (especially the reference to 'structural impartiality'), rather than the case by case consideration indicated in the *McGonnell* case. It is also difficult, if not impossible, to reconcile with Sir John Laws' approach in *McGonnell*.

If Article 6(1) does require a strict separation of powers as *Procola* might suggest, then the Law Lords could not continue to be members of the House of Lords; nor could the Lord Chancellor continue his mixed role.⁴⁴ If the less stringent approach to Article 6(1) is applied to the Law Lords' position then a more nuanced approach is required.⁴⁵ Broadly, one would have to consider whether each of the

39 The Commission notes that the then forthcoming decision in the case of *McGonnell v United Kingdom*, *The Times*, 22 February 2000, may have relevance to the position of Lord Chancellor. *McGonnell v United Kingdom* was a challenge based on Art 6(1) to a decision to refuse planning permission to convert a packing shed into a house. McGonnell's argument was that the Bailiff of Guernsey, by both presiding over the hearing of his planning appeal in 1995, as well as presiding (as Deputy-Bailiff in 1990) over passage of the island's development plan (on which the decision to refuse McGonnell's application had been based) breached Art 6(1) of the European Convention on Human Rights. See R. Cornes, 'McGonnell v United Kingdom and the Law Lords' [2000] PL 166.

40 *Procola v Luxembourg* (1995) 22 EHRR 193. In *Procola* the Court considered the functions of the Judicial Committee of the Luxembourg Conseil d'Etat. Four out of five members who sat in an administrative law matter had previously given a pre-legislative opinion on the instrument which was the subject of the administrative law matter. The Court held that this involvement meant that the Judicial Committee was neither objectively nor subjectively (a more difficult test to meet) impartial.

41 Concurrence of Sir John Laws, *ad hoc* judge in *McGonnell v United Kingdom*, n 39 above.

42 *ibid*, para 55 emphasis added

43 n 40 above, para 45.

44 While deciding that the Law Lords' mixed legislative and judicial roles were within its terms of reference the Commission baldly stated the Lord Chancellor's combined judicial, legislative and executive roles were outside its terms of reference and a matter for the Prime Minister (para 9.11). Neither the Law Lords nor the Lord Chancellor are specifically referred to in the terms of reference.

45 See further Cornes, n 39 above, 174–176.

various roles a Law Lord may carry out (from simply sitting in the upper house without speaking or voting, through to chairing committees, or actually speaking and voting in debates) might in individual cases amount to a breach of Article 6(1). The mere presence of a Law Lord in the chamber during discussion of a piece of legislation would be far less likely to cause an Article 6(1) concern than if a Law Lord spoke or voted on such a Bill.

The Royal Commission recommended that the Law Lords should 'publish a statement of the principles which they intend to observe when participating in debates and votes in the second chamber and when considering their eligibility to sit on related cases' (Recommendation 59). Their Lordships have now acted upon this recommendation.⁴⁶ The statement follows the majority position in *McGonnell* and also adopts the statement of the law in the *Locabail* case.⁴⁷ With the Law Lords soon being faced with Human Rights Act 1998 and devolution cases, it will become increasingly important that they have not been involved in the passage of, for example, a Bill subsequently challenged under the Act, or an apparently anodyne piece of Westminster legislation which is subsequently argued to be inconsistent with legislation from a devolved legislature.⁴⁸ The only sure way of avoiding an Article 6 (1) problem remains the removal of the Law Lords from the legislature and the removal of the Lord Chancellor from the bench.

Appointment of Law Lords

The Commission recommended that the terms for appointment of the Law Lords be brought into line with other members of the chamber, subject to automatic reappointment for as long as they are entitled to sit on the Appellate or Appeal Committees (Recommendation 58). That is, Law Lords would *prima facie* be appointed for limited 15 year terms. In its recommendations on an Appointments Commission (see 'Composition' below) the Royal Commission states that approval by the Appointments Commission should be the only way to gain entry to the chamber so that: 'the differences between members of the second chamber ... be minimised and the ability of the chamber to act as a cohesive body [is] enhanced' (para 13.4). To ameliorate the Commission's stated desire that differences between, for example, the Lords of Appeal in Ordinary and let us say, former MPs made Lords, be minimised they recommend that the Appointments Commission should have no discretion over whether or not to appoint those nominated to be Law Lords (Recommendation 82). The nomination process would continue to be the current process run by the Lord Chancellor.

The Commission's proposals on appointment of Law Lords are framed from the point of view of reforming the chamber as a legislative body. From the point of view of how members of a court, let alone the highest court, should be appointed, the recommendations fail to engage with the problems and concerns relevant to the

46 HL Deb vol 614 col 419 22 June 2000. The extent to which other Lords who may sit on appeals will observe these principles is not indicated. The Lord Chancellor has indicated he would also observe the statement of the Lords of Appeal in Ordinary HL Deb vol 614 WA col 163 7 July 2000. When making that statement he also referred to previous answers he had given in the House on the judicial position of the Lord Chancellor HL Deb vol 597 col 710-739 17 February 1999; HL Deb vol 602 col 1060-1063 24 June 1999; HL Deb vol 611 WA col 33 23 February 2000, HL Deb vol 610 col 655-657 2 March 2000

47 *Locabail (UK) Ltd v Bayfield Properties Ltd* [2000] 1 All ER 65.

48 The fact that almost all the Law Lords took part in some way in passage of the Human Rights Act, and that both Scottish Law Lords took part in passage of the Scotland Act at least raises the possibility that they could be challenged in relation to hearing subsequent cases under those Acts.

selection of judges (as opposed to selecting legislators). While there may be criticisms to be made of the current appointments process,⁴⁹ if the judicial appointment process were to be reformed this should be done on the basis that the new arrangements are qualitatively preferable in principle to the current arrangements. Simply proposing that the appointment power for the highest court's judges be given to the body responsible for appointing members of the country's legislature itself suggests the oddity of the Commission's recommendations. It is not clear on what basis the Commission thought that mixing the current process with a new and entirely untested one would avoid having an adverse impact on what they describe as the 'established independence' of the current process (para 13.6).

The composition of the chamber

Critiques of the Royal Commission report in the press have tended to focus on the proposed composition of the chamber rather than its powers and functions. As identified above, the Commission propose some important responsibilities for the upper house, although these fall short in some areas. It will be expected to scrutinise legislation in detail, with a potential to delay it for a year, to scrutinise government, carry out committee enquiries, and take on new roles regarding human rights and the constitution. For all of these tasks it is important that the chamber has a membership which is fit for its purpose, and which can command respect. Just like the hereditary peerage previously, there are growing concerns about the current method of appointment to the chamber and levels of political patronage. The Royal Commission proposes to replace this system with a chamber largely comprising two elements: appointees selected by a new Appointments Commission, and elected members representing the nations and regions of the UK.

A new Appointments Commission

Under the Royal Commission's proposals, the majority of members of the upper house would continue to be appointed, but with a statutory Appointments Commission taking over the appointments role currently performed by the Prime Minister. The Commission would be required to maintain a balance between the parties which reflected the votes cast at the last general election, but would also be required to ensure that at least 20 per cent of members were not aligned to any of the three main parties. It would have a statutory duty to ensure that at least 30 per cent of upper house members were women, rising to parity, and would ensure that the chamber also contained a fair ethnic and regional mix. The Appointments Commission would publish a regular statement saying what characteristics are sought in new members and would advertise vacancies and actively solicit nominations. It would keep the membership of the chamber under review and would report annually to parliament.

Under the Royal Commission's proposals the power of patronage would be removed entirely from the political parties. Parties would be free to nominate, but

49 Recent critiques include K. Malleon and F. Banda, *Factors affecting the decision to apply for silk and judicial office* (Lord Chancellor's Department Research Paper 2/2000); K. Malleon, 'The Peach Report on Silk and Judicial Appointments' (2000) *New Law Journal* 150; comments by Lord Steyn, 'Law Lord calls for democratic system to appoint new judges' *The Independent* 12 July 2000.

not to determine who sat on their benches – the Appointments Commission could select party representatives who had been nominated by other organisations. The removal of patronage power from party leaders could go some considerable way towards improving public confidence in non-elected members of the house. The requirements on party political balance would also ensure that no party had an overall majority in the chamber, and that the balance of power rested – as it does now – with minor parties and independents.

One of the Royal Commission's main justifications for including appointed members in the chamber was that 'people with relevant experience of and expertise in other walks of life [than politics] . . . would generally be reluctant to commit themselves to a party platform or engage in electioneering' (para 11.9). This is certainly true, particularly if the chamber is to include members who do not take any party whip. These members make up an important part of the chamber at present, marking it apart from most parliamentary chambers worldwide. The Royal Commission also took steps to boost the independence of party representatives from their respective whips. Members would serve 15 years in the chamber, with one renewal possible. Members would be barred from standing for the House of Commons for 10 years after their term expires, which would reduce the risk of members behaving as party careerists. However, ministers would continue to be appointed from the chamber, as discussed above. The proposals for an Appointments Commission thus seek to build on the British tradition of an appointed and 'expert' house, whilst boosting public confidence in the appointments system. The proposals appear to be effective in this regard. However, the proposed balance between appointed and elected members is a matter for concern. This is discussed in the following sections.

Elected regional members

The Royal Commission's proposals would also include a minority of elected members in the reformed upper house. The Commissioners were unable to agree what proportion of members should be elected, and therefore put forward three alternative models. Under each model the elected members would represent the nations and regions of the UK, with electoral boundaries in England using Regional Development Agency areas (also now used for European Parliamentary elections). All models would use a form of proportional representation, with a number of seats per region based on population. Two would run elections alongside European Parliament elections and one alongside general elections. In each case, upper house members would serve the equivalent of three parliamentary terms but with one third of members elected each time. Members would thus serve 12–15 year terms. Under the three schemes elected members would comprise 65, 86 or 195 members in a house of around 550. This would represent 12 per cent, 16 per cent or 35 per cent of upper house members, respectively. The elected element in the chamber is the Royal Commission's main concession to devolution. It would be the role of elected 'regional' members to provide a link to their nations and regions and represent territorial views. However, as noted above, there would be no formal mechanisms provided for them to do so.

There are a number of technical difficulties with the three options put forward by the Commission. Under the 'minimalist' position, with just 65 elected members, seats would be allocated to parties on the basis of House of Commons votes, with no direct opportunity for voters to express a view in an upper house election. This

unprecedented system would lead to a range of problems.⁵⁰ In addition, under both this system and that with 87 members, each region would get to select new upper house members at only every third election. Only the 'maximalist' position, with 195 elected members, would result in enough members being elected in each region to provide a proportional result in all areas every five years. Unless voters were to be disenfranchised from upper house elections for 15 year spells, this would be the minimum number of elected members that would need to be included in the chamber.

Balance between elected and appointed members

The primary difficulty with the Royal Commission's proposals on composition is the suggested balance between elected and appointed members. Whilst there are arguments for the inclusion of both groups within the house, the Commission does little to justify the dominance of the chamber by appointed members, the majority of whom would continue to take a party whip. The Commission acknowledged that there is considerable public support for a chamber which is wholly or largely elected.⁵¹ However, they rejected a wholly elected chamber on a number of grounds. One important consideration was that such a chamber would be dominated by the parties and unlikely to include many independent members or experts. This would change the nature of the house and be a considerable loss. However, this difficulty could be dealt with by inclusion of a significant *minority* of appointed members, with an emphasis on appointment of independents. The Commission's other main justification for rejecting an elected chamber was that the legitimacy afforded by election could result in a challenge to the 'pre-eminence' of the House of Commons. We thus return to the discussion of the interpretation of this requirement in the Commission's terms of reference, and whether this has been provided by the formal statutory restrictions on powers which the Commission proposed.

The Commission's concerns illustrate the inextricable nature of the issues of power and composition of the upper house. In the twentieth century the House of Lords proved to be relatively weak in opposing the House of Commons. This was not because of the chamber's powers, which are quite formidable over ordinary legislation, but because its anachronistic membership made it difficult to justify challenging the elected lower house. Despite the Lords' power to delay legislation for a year, potentially forcing government to proceed with support of the Commons alone, the Parliament Acts have run their full course on only three occasions since 1911.⁵² On each occasion when a potential challenge occurred, the audacity of the unelected house was presented as presaging a constitutional crisis. During debate on the disputed European Parliamentary Elections Bill in 1998, for example, Home Secretary Jack Straw stated that '[t]he issue has now gone way beyond whether we have closed lists, open lists or Belgian lists: there is now a challenge to the authority of this democratically elected House by hereditary Conservative peers'.⁵³

50 For a discussion of these see M. Russell and R. Hazell, *Commentary on the Wakeham Report on Reform of the House of Lords* (London: Constitution Unit, 2000)

51 A poll conducted by the Democratic Audit whilst the Commission was sitting found that 84 per cent of the public favoured election over appointment for the upper house *The Guardian* 17 July 1999.

52 The 1911 Parliament Act was used to force through the Parliament Act 1949. The terms of the Acts were then used subsequently over the War Crimes Act 1991 and the European Parliamentary Elections Act 1999. For a discussion of the former see Shell, n 12 above, for the latter see G. Ganz, 'European Parliamentary Elections Act 1999: The Constitutional Issues' (1999) 18 *Amicus Curiae* 4.

53 HC Deb vol 337 col 206 10 November 1998.

How this situation would be changed by the establishment of a largely appointed chamber is difficult to predict. There is little that overseas experience can tell us. The Canadian Senate is wholly appointed, and the only chamber of its kind amongst Western democracies outside the UK. It has historically suffered from the same difficulties as the House of Lords, since its membership is seen as a hangover from a bygone age, long discredited by the blatant political patronage exercised by Prime Ministers in making appointments.⁵⁴ The Royal Commission's proposals for an Appointments Commission would avoid some of the aspects which have been most damaging in Canada. However, it is questionable whether the reputation of a largely appointed chamber could survive the battering it would be likely to receive from successive governments, if the will of such governments were frustrated. It is notable that when the unreformed House of Lords attempted to amend the Welfare Reform and Pensions Bill in 1999, Social Security Secretary Alistair Darling referred to the interference as 'a constitutional matter', saying that 'the House of Lords has to accept that we are the elected chamber'. There is reason to believe that such criticisms would continue for as long as the upper chamber remains unelected.

The function of the upper house proposed by Bryce of 'interposition of delay' remains one of its most important potential roles today. Whilst the Royal Commission propose to give the reformed chamber sufficient formal powers to perform this essential function, it is questionable whether the chamber would have sufficient legitimacy to fulfil it. The Royal Commission recognised the importance of legitimacy, stating that 'Throughout the 20th century the House of Lords was inhibited both by its lack of authority and its lack of confidence. The reformed second chamber must be free of such debilitating inhibitions' (para 10.8). In justifying the proposed mix between elected and appointed members the report states that '[a] second chamber with at least a significant proportion of directly elected members would have the necessary political weight to carry out the responsibilities we propose it should have and its decision would be more widely seen as politically legitimate' (para 11.3). The question is whether 12 per cent, 16 per cent, or even 35 per cent of members is a sufficiently 'significant proportion' for this to be true. Earlier in the report the Commission also states that there is 'in our view, not much point in the second chamber having a theoretically greater power which it does not in reality exercise. It should have powers which it can actually exercise, and which would require the Government and the House of Commons to take some positive action' (para 7.35). In providing the new chamber with a largely appointed membership it seems unlikely that the Royal Commission will have met this requirement.

Another difficulty arising from the proposed balance between elected and appointed members is that it undermines the Commission's proposal that the chamber should take on an important territorial role. Only a minority of members would be selected by the nations and regions themselves. Although the Appointments Commission would be required to ensure a geographical balance amongst all members of the chamber, the Commission's report itself states that 'we were told at our public hearing in Newcastle, people in the regions would not regard someone selected *for* their region by a London-based Appointments Commission as being an adequate substitute for someone selected *by* their region' (para 11.31). This is another weakness of the Senate of Canada, where members

54 See Russell, n 8 above, 225–226; C. E. S. Franks, 'Not Dead Yet, But Should It Be Resurrected? The Canadian Senate' in Patterson and Mughan (eds), n 32 above.

are appointed to represent their provinces, but these appointments are made by national government. As a result the Senate is not seen as a proper representation of the provinces. This appears to be a real danger with the Royal Commission's proposals.

There are few precedents internationally for a mixed elected/appointed upper house. The only chambers to include a significant proportion of both elected and appointed members are those in Malaysia, Swaziland and Algeria. However, the concept of a mixed chamber has long been on the British political agenda. In 1978 the Home Committee established by Margaret Thatcher proposed an upper chamber where one third of members were appointed and two thirds were elected.⁵⁵ This mix was favoured by Lord Richard, Labour Leader of the House of Lords from 1997–98.⁵⁶ Whilst the Liberal Democrats proposed an all-elected chamber to the Royal Commission, the Conservative Party proposed two models, one of which would be a mixed chamber where 86 per cent of members would be directly elected. The Royal Commission stood almost alone therefore in proposing that elected members should make up a *minority* of members of the house, although the government have now accepted that principle.⁵⁷

Conclusion

We conclude, therefore, that overall the Royal Commission's report is timid. Contrary to its terms of reference, the Commission has not taken sufficient account of the new constitutional settlement which is now bedding down in the UK. The Commission has also not considered broadly enough the appropriate powers for an upper house in a modern British democracy. These are perhaps inevitable failures given the pressures under which the Commission was required to report. The advent of devolution and the incorporation of the ECHR will have profound effects which are as yet difficult to quantify. At the centre of a new British constitution there is a place for an upper house with clear territorial functions, with constitutional guardianship powers, and with a composition that enjoys sufficient democratic legitimacy to enable the chamber to play a full part in our system of government. Given our current constitutional arrangements, the reform of the upper house also requires an examination of the relationship between the judiciary, executive and parliament. The Royal Commission's report does not provide this – and it is difficult to see how it could, given the constraints within which the Commission operated. Far from proposing a 'House for the Future', the report's approach is too gradualist meet the demands of what is now a radically different Britain. With the hereditary peers largely removed from the chamber, the debate has already moved on. The Commission's report provides a starting point, but probably no more.

55 Lord Home, *Report of the Review Committee on the Second Chamber* (London: Conservative Political Centre, 1978).

56 Lord Richard argues his position in I. Richard and D. Welfare, *Unfinished Business: Reforming the House of Lords* (London: Vintage, 1999).

57 When the Royal Commission's proposals were debated in the House of Lords Baroness Jay, the Leader of the House, welcomed the report and indicated that its principles were accepted. In particular she stated that the new chamber should be 'largely nominated but with a minority elected element and with a particular responsibility to represent the regions' HL Deb vol 610 col 912 7 March 2000.

CASES

Increasing the Price of Pain: Damages, The Law Commission and *Heil v Rankin*

*Richard Lewis**

Five judges in a specially constituted Court of Appeal have decided in a conjoined appeal, involving eight test cases,¹ that damages for pain, suffering and loss of amenity (PSLA)² must be increased. Claims for non-financial loss must be revalued. But the changes *Heil v Rankin*³ makes are by no means as radical as some feared, and others hoped. This is because, firstly, the case only affects the small minority of personal injury claims which involve serious injury; and secondly, even in the most extreme of these cases, it increases damages only by a third, and in most of the affected cases the increase will be a great deal lower than this. The overall effect is to boost damages by much less than was wanted either by the influential claimants' lobby⁴ or by an 'impertinent'⁵ Law Commission which recently reported on the matter. The extra money now demanded of insurers can be found relatively easily. The case will remain of importance for some time to come for both sides of personal injury practice feel that they have reached a stalemate in this particular battle. No further litigation is now planned.⁶

According to the court, the reasons for the increase are that it is 'fair, reasonable and just,' it accords with the view of society as a whole, and it results in 'proportionate' awards which take into account not only changes in economic circumstances but also the impact of awards upon society. Future levels are to be adjusted in line with the rise in the Retail Prices Index, and further appellate consideration will not be necessary unless the awards again become significantly

* Cardiff Law School. I am grateful to several colleagues at Cardiff for their comments on an earlier draft of this article and, in particular, to David Campbell, Ken Oliphant and Richard James.

- 1 Although there was some difficulty in gathering together representative cases, eventually they included, at the top end of the scale, children suffering from cerebral palsy, and at the bottom end of the scale, whiplash injuries. Appearances were made by 12 juniors and an equal number of silks, including one who appeared as an *amicus curiae* instructed by the Treasury Solicitor. Written submissions were received from the Association of Personal Injury Lawyers, the Forum of Insurance Lawyers, the Association of British Insurers, the Eagle Star Insurance company, and the Iron Trades Insurance Company.
- 2 This abbreviation, used consistently in *Heil*, is less familiar to academics than the term non-pecuniary loss. Practitioners frequently use the term general damages to mean PSLA even though, strictly, this includes all items of loss which cannot be specially pleaded, and therefore includes future financial loss. Although many may regret the increasing use of acronyms, at least there might be less misunderstanding between academics and practitioners if PSLA is commonly adopted.
- 3 [2000] 3 All ER 138, [2000] 2 WLR 1173, [2000] PIQR Q187.
- 4 See the Association of Personal Injury Lawyers response to the Law Commission Consultation Paper No 140 on non-pecuniary loss, and its submission made to the Court of Appeal in *Heil's* case.
- 5 The Law Commission Report No 257, *Damages For Non-Pecuniary Loss* (1999) is so described by Tony Weir, *A Casebook On Tort* 9th ed (London: Sweet and Maxwell, 2000) 636.
- 6 Although none of the claimants applied for permission to appeal when the Court of Appeal judgment was handed down, two of them later did seek leave from the House of Lords. However, from information supplied to the author by the relevant firm of solicitors, Russell, Jones and Walker, the appeal is not now being pursued because of concerns about the costs and the risks involved.

out of line with the standards identified. To the casual observer, these views appear eminently sensible. However, the judgment and, more particularly, the earlier Law Commission report leave this reviewer with a sense of unease. Certain perspectives upon the tort system, and the criticisms which have been made of it, seem to have been entirely forgotten. This note, in a polemical fashion, tries to restore a little of the balance.

It does so by emphasising the disproportionate importance of PSLA in the award of damages, and the failure of students and others to appreciate what the tort system in practice actually achieves. It notes that PSLA is a major cause of the excessive cost, inefficiency and injustice of the tort system. It casts doubt upon surveys purporting to justify an increase in damages for pain and suffering by reference to the popular will. It reminds readers of criticisms of tort based on distributive justice arguments, and it contrasts these with the corrective justice perspective which determined not only the scope of the Law Commission review of the law of damages, but also the proposals for reform it put forward. It concludes that the decision in *Heil v Rankin* is not as obviously in the public interest as the rhetoric of the judges or the Law Commission would have us suppose.

Calculating the limited increase in damages

The Law Commission proposed a large increase in damages for PSLA. A wide range of cases would have been affected. It suggested that awards over £3,000 should be doubled or increased at least by half.⁷ By contrast, *Heil v Rankin* makes no change in damages for non-pecuniary loss assessed at, or below, £10,000;⁸ but where such damages would have been £150,000, as in the case of tetraplegia or severe brain damage, the amount is to be increased by a third to £200,000; and between £10,000 and £150,000 there is to be a progressive increase from 0 per cent to 33½ per cent so that, for example, in the middle of the range, an award of £80,000 is to be increased by 17 per cent to about £95,000.

At the end of its judgment the court even included a simple graph indicating the approximate level of increase. This unique appendage to a law report can make the calculation look easier than in fact it is, and a mathematical formula has since been published to enable the new level of damages to be set.⁹ Lawyers traditionally have been amongst the least numerate of professionals, and suspicious of anything other than words within which to encompass ideas and rules.¹⁰ However, in recent years the increasing ability of personal injury practitioners and judges to understand and use the information provided to them by accountants and actuaries has fuelled important developments in tort law. For example, spread sheets and discount rates are not quite the mysteries they once were, and structured settlements have broken the mould of awarding damages only in the form of a lump sum.¹¹ In the same financial vein, litigation based on the real rate of return

7 Report, above n 5, at para 56. For injuries between £2,001 and £3,000 it recommended a tapered increase of less than a factor of 1.5 so that an award of £2,500 should be uplifted by 25 per cent.

8 This threshold was suggested by the Forum of Insurance Lawyers at p 45 of their submission to the court. It would cut out of the system almost all claims presently processed under the "fast track," being for amounts below £15,000 in total, including financial loss.

9 Somewhat surprisingly, the author could find no other instance of the use of a graph in a UK judgment. For the mathematical formula see Levene and Vaughan, 'Making *Heil v Rankin* Work' [2000] *Quantum* (April 18).

10 This is so even in tax cases. See eg Beldam LJ in *Smith v Schofield* [1992] STC 249.

11 R. Lewis, *Structured Settlements: The Law And Practice* (London: Sweet and Maxwell, 1993).

from investment in equities as opposed to gilts has been a major focus of concern.¹² The graph used in *Heil v Rankin* to update awards compared with past values is thus symbolic of the closer relationship being drawn between personal injury law and the wider financial world.

Within four months of the judgment the increase in damages was reflected in the new edition of the Judicial Studies Board guidance on assessing damages for non-pecuniary loss,¹³ a book to be 'packed in every judge's lunchbag'.¹⁴ We may not have a statutory cap on damages as exists in parts of the USA¹⁵ and Australia,¹⁶ nor do we yet have a formal table of compensation based upon a percentage assessment of disablement as is being developed within the European Union,¹⁷ but the informal brackets within which PSLA damages are to be assessed are becoming increasingly detailed. They provide further ammunition for an attack on the rhetoric of tort law. It is increasingly difficult to assert that a subjective individualised system of justice is being employed to return the claimant to anything even approaching the pre-accident position. No matter that it is the index finger of a concert pianist that has been lost, she will get only a little more than the £9,500 PSLA specified in the book. Of course, she may also be entitled to a large sum for loss of earnings, but the judge will still be incapable of performing a digital transplant. PSLA is compensation in a very special sense; unlike damages for money lost and expenses incurred, it cannot provide an equivalent to that which has been destroyed. It is an artificial sum and, increasingly, it is mechanically assessed.

The importance of small claims to the tort system

Although both the Court of Appeal and the Law Commission looked at the wider social and economic circumstances which related to the proposed increase in damages, there is one crucial fact which they either ignored or which escaped their attention. It is the foundation for a perspective upon personal injury claims which is of vital importance, and it is extraordinary that the Law Commission, in particular, failed to highlight this fact. Buried in the middle of its report is an acknowledgement that two thirds of the total damages paid for personal injury are

12 *Wells v Wells* [1999] AC 345, [1998] PIQR Q56

13 Judicial Studies Board, *Guidelines For The Assessment Of General Damages In Personal Injury Cases* 5th ed (London: Blackstone Press, 2000).

14 Weir, above n 5, at 637.

15 For a review of the imposition of such limits on damages in the USA see Sanders and Joyce, 'Off To The Races: The 1980's Tort Crisis And The Law Reform Process' (1990) 27 Houston LR 207. See also the Law Commission Consultation Paper No 140, *Damages For Personal Injury: Non-Pecuniary Loss* (1995) paras 3 62-3.64

16 See Mullany, 'A New Approach To Compensation For Non-Pecuniary Loss In Australia' (1990) 17 Melbourne Univ LR 714, and the Law Commission Consultation Paper, above n 15, paras 3.29-3.32. By contrast in Canada a judicial rather than legislative limit of \$100,000 was imposed by the Supreme Court in a trilogy of cases, the most notable being *Andrews v Grand Toy Alberta Ltd* [1978] 2 SCR 229. The limit was later increased to take account of inflation so that by 1999 a ceiling in the region of \$265,000 had been established. See J. Cassels, *Remedies: The Law Of Damages* (Toronto: Irwin Law, 2000) 159. The limit is comparable to that specified by the JSB guidelines.

17 The author was one of few British representatives at a congress dealing with the rationalisation of medico-legal assessment of non-economic loss which was held in Germany, at the Academy of European Law at Trier, in June 2000. Chaired by a Member of the European Parliament, the congress received a report and drafted a proposal for an addition to the Fourth Directive relating to civil liability and motor insurance. Its objective is to produce 'harmonised compensation' by adopting 'a compensation scale, expressed in points, which would enable every Member State to assign a monetary value to the results of a medical assessment.' cf the Law Commission's abrupt dismissal of medical scoring in its 1999 Report, above n 5, at para 3.194.

for non-pecuniary loss.¹⁸ Much can be made of this, together with the Commission's failure to cost its proposals for change.

Students of tort traditionally focus excessively upon liability issues, and spend little time examining damages. As a result they often completely misunderstand what the system in practice achieves. The 'missing' fact reveals that the main function of the tort system is not to provide for the future loss of income and care needs of those seriously disabled by accident or disease.¹⁹ Such especially needy claimants are relatively rare. Instead the system overwhelmingly deals with small claims, the great majority leading to damages of less than £5,000.²⁰ In these cases claimants suffer very little, if any, financial loss. They make a full recovery from their bodily injury and have no continuing ill effects.²¹ They make no claim for any social security benefit as a result of their accident.²² In the minority of cases where benefit is claimed, PSLA is the only part of the award of damages from which no social security can be deducted.²³ This means that in a few cases the damages claim, in effect, is being made only for the non-pecuniary loss. In settlements in general the largest component by far is the payment for pain and suffering. The stereotypical injury is the minor whiplash which follows a low-speed car 'shunt.' It is these type of cases which account for the extraordinarily high costs of the system compared to the damages it pays out.²⁴ If it

18 As found by the *Report Of The Royal Commission On Civil Liability And Compensation For Personal Injury* (1978, cmd 7054), chairman Lord Pearson, vol 2 table 107 (the Pearson report). Although the Law Commission notes the fact at para 3.38 of its report, it fails to give it the prominence it deserves.

The relative importance of PSLA is reduced by two factors. First, as revealed by the Pearson report table 108, the larger the claim, the lower the proportion that is paid for non-pecuniary loss. For claims over £25,000 in 1973 the proportion paid for non-pecuniary loss dropped to 48 per cent. This remains an extraordinarily high figure. Secondly, in more recent years it is likely that the proportion of PSLA has reduced because there has been an increase in the level of damages for pecuniary loss. However, there are no figures to illustrate this. Counteracting these two factors are other developments which have increased the importance of PSLA. For example, the effect of the Social Security (Recovery of Benefits) Act 1997 is discussed in n 23 below.

19 Future pecuniary loss was found in only 7.5 per cent of cases, and comprised only 8.3 per cent of the total damages paid in tort. See the Pearson report, above n 18, vol 2 para 44 and table 107.

20 In 1995 the DSS estimated that half of the cases reported to it under the compensation recovery scheme were being settled for £2,500 or less. DSS *Memorandum of Evidence to the Social Security Select Committee* (1995) HC 196, para 40. of the figure of 37 per cent given 5 years earlier in HC Deb vol 166 col 942 February 7 1990. £2,500 was the median figure in the survey of 81,000 cases receiving legal aid and closed in 1996-97 in P. Pleasence, *Report Of The Case Profiling Study, Personal Injury Litigation In Practice, Research Paper 3* Legal Aid Board Research Unit (1998), 40 fig 3.17. Similarly in evidence to the Law Commission in March 1993 the Trades Union Council noted that the average sum obtained in the 150,000 union-backed cases in 1991 was under £2,000. See also the figures for various years in D.R. Harris et al, *Compensation For Illness And Injury* (Oxford, Clarendon Press, 1984) 87.

21 Pleasence, n 20 above, found that about three quarters of all successful claimants had fully recovered three years after their accident. Pearson, above n 18, vol 2 table 70 noted that 95 per cent of accident victims were back at work within six months of their injury. But Genn found that 39 per cent of victims suffering accidents in the last six years reported that they were in constant pain. See Law Commission Report No 225 *Personal Injury Compensation: How Much Is Enough?* (1994), 58.

22 The Compensation Recovery Unit issues a nil certificate in 70 per cent of the cases where a compensator intends to make a payment. See R. Lewis, *Deducting Benefits From Damages For Personal Injury* (Oxford, OUP, 1999) para 14.05.

23 Lewis, above n 22, para 13.33. As a result of the reforms made by the Social Security (Recovery of Benefits) Act 1997 the compensation awarded for PSLA has now been ringfenced so that no deduction for benefit can be made from that part of the award.

24 The figures always bear repetition. The Pearson Commission, above n 18, estimated that the cost of operating the tort system amounted to 85 per cent of the value of tort payments distributed to claimants. See vol 1 para 256. The Lord Chancellor's *Civil Justice Review* (Cm 394, 1986) estimated that the cost of the tort system consumed 50 to 70 per cent of the total compensation awarded in personal injury cases. See also Pleasence, above n 20, 64, who found a damages to costs ratio of 2.27:1, and costs exceeding damages in 22 per cent of successful cases.

had been wholly in public hands, tort would have been radically revised and then privatised some years ago.

The 'missing' fact is thus of crucial importance in understanding the tort system, especially when its future is being debated and the relative priority of competing claims for compensation is being considered. It is also fundamental in making an assessment of the cost of any changes to be made. The extraordinary failure to focus on the relative importance of PSLA is perhaps a measure of how attitudes have changed since the post-Thalidomide era of the 1970's when the role and future of the tort system as a whole was in question. Although our empirical understanding of the tort system has continued to improve, academics and law reformers are now less likely to address the fundamental issues exposed by the facts and figures.

Small claims thus constitute the everyday battleground over which insurers and claimants' representatives fight. Their predominance is the reason why insurers have not been too dismayed by *Heil v Rankin*. Not only is there to be no increase in PSLA assessed at £10,000 or below, but also the tapered increase above that threshold figure is such that, at the lower end of the scale, the effect of the increase is negligible. Even at £30,000 – corresponding to the total loss of an eye – the increase in damages is less than £1,500. This is not enough to upset the course of settlement negotiations in the vast majority of cases.

Defendants were fearful of damages being increased in minor cases along the lines suggested by the Law Commission. According to the Association of British Insurers, the Commission's proposals would have cost insurers between £500 and £1,000 million more per year.²⁵ That would be ten per cent of their total premium income for motor, employers liability and public liability, estimated as about £10 billion in 1999. Whereas it would have added 27 per cent to the cost of employers liability claims, only 8 per cent would have been added to the cost of motor vehicle insurance²⁶ (which predominantly reflects property damage costs rather than personal injury). About 168,000 claimants a year – perhaps a quarter of the total²⁷ – would have had their damages increased. With the much higher threshold set by *Heil v Rankin* at £10,000 only a few thousand claimants a year will now benefit, and the increase in their damages will be much less.

Justifying the increase

Which forum?

The Court of Appeal considered whether it was the appropriate forum for responding to the Law Commission's invitation to reconsider the level of damages,

25 See Bacon and Woodrow, "Assessment Of The Possible Impact Of Reform Of The Law Of Damages For Pain And Suffering On UK Insurers' Claims Costs" (February 2000), a report submitted to the Court of Appeal by the ABI

26 *ibid.*

27 In the six months to the end of September 1998 the Compensation Recovery Unit was receiving notification of about 54,000 claims being made against compensators each month. R. Lewis, *Deducting Benefits From Damages For Personal Injury* (Oxford: OUP, 1999) para 14.07. This is an annual rate of about 650,000 a year. Thus the ABI figure of 168,000 claimants affected is roughly a quarter of the total. However, this does not take account of the fact that the annual figure includes all tort claims, not just those where the compensation was paid by a member of the ABI. In addition, these figures must now be revised upwards because the numbers of claims has continued to grow. According to information received by the author from the DSS, the total number of claims notified in the financial year 1999–2000 was over 716,000, a rise of ten per cent since the six months to the end of September 1998.

or whether the matter should be left to Parliament. It eventually decided that it was appropriate for the court to do so because it was doing no more than applying existing legal principles and updating them. It made what is becoming a ritualistic reference to the European Convention on Human Rights and rejected the suggestion that a retrospective increase in damages might fall foul of the Convention. The court again signified that it would not allow the Convention to dominate matters. It robustly stated that it was justified to hand down a judgment with retrospective effect because the possibility of change is one which insurers and others ought to foresee,²⁸ and the increase in damages was in the public interest because it provided fair, reasonable and just compensation.

The influence of surveys and representations

The court decided that it was only appropriate to interfere with the level of damages if there was a clear need to do so. Although persuaded by the Law Commission report that there was such a need, it took a different view of the material upon which the Commission relied. In particular, it attached much less importance than the Commission to the surveys and empirical studies which had formed an important feature of its report, although the court emphasised that the level of damages must not be out of line with what society as a whole would view as being reasonable. The Commission relied heavily on four sources in order to justify increasing damages in more serious cases:

- A study it commissioned from the Office of National Statistics (ONS) concerning public opinion about the level of damages.²⁹
- Hazel Genn's 1994 general report for the Commission's damages project entitled 'How Much Is Enough?'³⁰
- The findings of the Legal Aid Board Research Unit into the settlement of 81,000 legally aided personal injury cases.³¹
- The responses the Commission received to its Consultation Paper.³²

The first two studies were much criticised by the defendants in *Heil*,³³ and the Court of Appeal preferred instead to rely on the response to the consultation. This had strongly endorsed an increase in the upper level of awards. Although half of the respondents agreed that damages for minor injuries were not too low, the Law Commission reported that 74.5 per cent of them believed that damages had failed to keep pace with inflation and that more should be awarded to those suffering very serious injuries. As usual, these respondents were predominantly lawyers closely connected to the present system, most of them being personal injury practitioners. The counting of heads and assembling of percentages in relation to those who reply to Law Commission documents is not likely to report results which reflect the views of other than a cross-section of interested lawyers. It is not surprising that they are likely to carry weight and be acceptable to judges. However, these

28 cf Bennett's concern about insurers' setting of premiums and access to reserves in 'Personal Injuries: General Damages – Minor Increases Only' [2000] JPIL 129 at 135.

29 Office Of National Statistics, *Omnibus Survey: Module 209 Compensation* (1998).

30 see n 21 above

31 P. Pleasence, *Report Of The Case Profiling Study, Personal Injury Litigation In Practice, Research Paper 3* Legal Aid Board Research Unit (1998).

32 n 15 above.

33 The ABI commissioned a report from James Rothwell, an independent marketing and economic research consultant, which made detailed criticisms of the adequacy of the research upon which reliance was placed.

percentage opinions by no means necessarily reflect whether society as a whole would perceive the increase in damages as reasonable, and yet this was the key factor which the court continued to stress.

When it came to considering the empirical surveys, the court recognised that there were great difficulties in carrying out such work. However, it severely criticised the ONS survey for failing to draw sufficient attention to the fact that very substantial damages for pecuniary loss could be awarded in addition to that for PSLA. Interviewees were not told, for example, that the provision of aids and equipment, adapted accommodation, and holidays would be specifically covered in the award for pecuniary loss. Nor were they necessarily aware that, in the most serious injury cases, the award for lost earnings and the cost of care is likely to be many times the size of that for PSLA.³⁴ Similarly, the dissatisfaction found by Genn in her survey of recipients of damages awards could be explained for reasons other than unhappiness about the amount given for PSLA.

With regard to the ONS attempt to seek views after trying to form a link between an increase in damages and the payment of insurance premiums, the court thought that the link had not been made sufficiently explicit and therefore the survey was unreliable. It also favoured the investigation of another link not pursued by the ONS. It would have liked to have seen what people would have said if a connection had been drawn between an increase in damages and the diminished resources which may then be available to the National Health Service (NHS) as a result.

The use of 'vox pop' to justify changes in the tort system is fraught with danger. The use of simple, basic questions can produce predictable responses. It would be no surprise to find that people would award more compensation to those with serious injuries, just as they would have 'the government' increase the wages of nurses, double the value of old aged pensions, and waive contributions to residential care. The difficulties lie, firstly, in trying to assess the extent to which they would pay to fund the increased payments once they realise it is they who are paying, and, secondly, in placing their proposals for increased expenditure in some order of priority. The formulation and scope of the questions depends upon the perspective from which the surveys are conducted. One major premise at the heart of the Law Commission series of reports on damages is that the tort system itself must remain in place. Its remit was to review the law of damages for personal injury but this did not allow for any wider review of the efficacy of the tort system as a whole. This crucial limitation is briefly mentioned at the beginning of a couple of its earlier reports.³⁵ In the middle of its report on PSLA the Commission notes that it could not recommend any trade-off with a new no-fault compensation scheme,³⁶ and later it acknowledges that its terms of reference prevent it from reviewing whether the tort system is justified on corrective justice grounds.³⁷ Indeed, the Law Commissioner most involved with the damages project has since written, in his private capacity, an extraordinary defence of the

34 Thus in one of *Heil* test cases, *Warren v The Northern General Hospital NHS Trust*, the Court of Appeal increased the damages by £40,000 but this amounted to less than 1.4 per cent of the total award of almost £3 million.

35 See eg Consultation Paper No 147, *Damages For Personal Injury: Collateral Benefits* (1997) para 1.7. 'in line with our terms of reference, our role in this review is not to advocate the replacement of the existing tort system but rather, assuming its continued existence, to recommend improvements to it.'

36 Para 2.26 of Law Commission Report No 257, n 5 above.

37 *ibid* para 3.60: 'we accept a corrective justice conception and justification for tort law. As such we believe that the tort system's damages rules should be developed in accordance with that theoretical perspective.'

tort system as morally justified because it pins legal responsibility upon an individual for his conduct and deters wrongful behaviour.³⁸ The continued existence and efficacy of the law of tort are explicitly accepted by the Commission, and reflected in the construction of the surveys upon which the Commission relies.

To illustrate, let us consider questions which people might have been asked if a wider perspective on tort had been taken. These questions explicitly deal with relative priorities for expenditure. The likely responses to these questions are no less predictable than those asking whether victims should get more money, but they carry very different implications for the future of damages for non-pecuniary loss. The first question begins by explaining that more than two-thirds of accident victims who are so seriously injured that they are unable to work for more than six months are unable to claim any damages at all and, instead, must rely on social security benefits.³⁹ The remaining third are able to claim not only for their full financial loss but also for their pain and suffering. The question then to ask is: 'When allocating further resources to accident victims, should more money be spent on those already able to claim damages by giving them more for their pain and suffering?'

Similarly, a second question, although taking a narrower focus, could still attack the priority now given to PSLA. It asks people to place in order of importance the losses which they might choose to insure themselves against if they were to take out a policy against being involved in an accident causing personal injury. Which would they regard as the most important loss to insure against: an interruption in earnings; the cost of medical and other care; or PSLA? The inevitably low priority given to PSLA would be all the more apparent if it were explained that the premium to be paid would be more than doubled if it were to cover pain and suffering. It is clear that those who are knowledgeable about the risks of accident or illness, and seek policies to protect themselves against such eventualities, do not wish to pay much higher premiums for a type of loss for which money cannot easily provide a substitute. If it were left to market forces there would be little cover for PSLA.⁴⁰ This is not to say that some form of recognition of non-pecuniary loss is not wanted in an ideal world, but it is clear that lost earnings and out-of-pocket expenses come first and, for the great majority of accident victims, these are losses which are not being met. The tort system simply fails to meet these basic needs. This view - which exposes the law of tort as both extremely limited in coverage and yet extravagant in content - cannot be found in the Law Commission report or the Court of Appeal decision.

The relevance of other courts and countries

Generally the Court of Appeal attached little importance to the value of awards in other jurisdictions and areas of law. For example, it gave less weight than did the Commission to the level of awards in Northern Ireland, which generally are double those found in the rest of the UK. It explained these higher awards as being linked

38 A. Burrows, 'In Defence Of Tort' in *Understanding The Law Of Obligations* (Oxford: Hart Publishing, 1998) especially pp. 122-126.

39 Pearson Report, n 18 above, vol 2 table 88.

40 This is conventional wisdom among economists. See Priest, 'The Current Insurance Crisis And Modern Tort Law' (1987) 96 Yale LJ 1521, at 1546-7. However, this is subject to attack from Croley and Hanson in 'The Non-Pecuniary Cost Of Accidents: Pain And Suffering Damages In Tort Law' (1995) 108 Harvard LR 1787.

to previous levels when juries had been used. On the other hand, the court was prepared to attach more importance than the Commission to awards in EU and EFTA countries. No reasons for this were given. Nor was there any justification advanced for dismissing the comparison with the level of criminal injuries awards, although this had been dealt with in detail by the Commission.⁴¹ Decisions of tribunals in employment and discrimination cases, and of courts in defamation and police liability cases were similarly thought irrelevant. Significantly, no mention was made by the court, and little by the Commission, of how compensation for the great majority of accident victims would be assessed: that is, no one noted the principles which a social security tribunal would employ.⁴² Of course, the difficulty with making comparisons is that for loss of enjoyment and injury to the senses there can never be a 'right' figure. Examination of alternative compensation can expose fundamental issues concerning the role of tort law and its relationship to other welfare systems, and the court gave little time to it.

The increase in life expectancy

In deciding to award more money only in cases of more serious injury, the court was much influenced by the increase in life expectancy which we now enjoy. It means that the effects of more serious injuries are endured for longer. The change is most apparent in cases of extreme injury. The life expectancy of paraplegics and tetraplegics, for example, is much greater than it was thirty years ago. Insurers countered the argument by noting that treatment and medication had also improved. Nevertheless, the court held that these fell far short of compensating the individual for the increased pain and suffering caused by permanent disabilities continuing throughout the increasing life span. Living longer as a collateral benefit was not discussed; instead longer life was equated with more pain and therefore more money.

Updating by reference to prices not earnings

According to Weir, the Law Commission report was unpersuasive: 'While stating that there is no demonstrably right view on quantum, its proposal to increase damages in line with inflation presupposes that the previous view was correct.'⁴³ But once a level is established, by what mechanism is it to be kept up to date? The court opted for a combination of the Judicial Studies Board guidelines and continued use of the Retail Prices Index to uprate past awards. Uprating by reference to tables relating to the rise in average earnings or to the rise per capita in the Gross Domestic Product was rejected. This was partly because the claimants had introduced these arguments at a late stage in the proceedings. Had the alternative indices been used, damages could have increased substantially. This is because, for example, between 1991–98 earnings increased by 30 per cent whereas prices rose only by 21 per cent.⁴⁴ However, it should be noted that the top end of

41 See paras 3.73–3.85 of Law Commission Report No 257, n 5 above.

42 The Commission dismissed Atiyah's criticism that its consultation paper had failed to examine social security payments. It boldly stated that they provided no useful guide because there was "no sense in which the social security system is designed to be corrective justice and therefore to be fully reparative." See n 5 above, para 3.71.

43 n 5 above, 636.

44 Over the longer period between 1970 and 1999 average earnings increased by 1433 per cent whereas the RPI only increased by 794 per cent. That is, average earnings increased by 1.8 times more than the RPI. Similarly, according to the ONS, between 1971 and 1997 real household disposable income nearly doubled.

the PSLA scale specified by the Judicial Studies Board exceeded this rise in prices for many types of injury. Atiyah long ago pointed out that, in relation to lost earnings, recipients of damages were frozen in time to the earnings levels applicable at the date of their settlements. They will not share the rise in the standard of living which will be enjoyed by the rest of us because of the anticipated increase in wages over prices (which has averaged two per cent a year since 1945). In that sense tort undercompensates accident victims. They have never benefited from the link Barbara Castle once forged between state pensions and earnings, a link later destroyed by the Thatcher administration.⁴⁵ The Court of Appeal decision to continue to use the familiar updating mechanism of the RPI was to be expected; any move away from it, even in the context of non-pecuniary loss, could have had wider implications.

The inter-relationship of heads of damage

Although *Heil* is an important case, it is but one of many which mark the struggle between insurers and increasingly well organised and experienced claimants' representatives. The Association of Personal Injury Lawyers is an extremely effective body, and outshines its opposing organisation, the Forum Of Insurance Lawyers, as a pressure group and information resource. Its success reflects the fact that damages in serious cases are radically different from what they were thirty years ago. During that time maximum awards overall have increased by between 40 and 60 times. The assessment of financial loss is now far more sophisticated and detailed. Although this was highlighted recently by *Wells v Wells*⁴⁶ and the debate about the appropriate discount rate to apply to arrive at the multiplier, there have been many other improvements in damages assessments which are now accepted as common features of tort law. In particular, claimants obtain compensation for items of pecuniary loss which once would have been subsumed under the general heading of loss of amenity. Two examples, were given by the Court of Appeal in *Heil*: claims for special holiday arrangements, and claims for adaptation and improvement of the conditions in which the claimant now lives. If specific sums are allocated for these purposes there is a danger of double accounting. For example, the inability to take a normal holiday could be taken into account as an item of financial loss as well as figuring in the overall assessment of PSLA. The court recognised this danger. In addition, losses which now were pleaded as aspects of the financial loss claim rather than forming part of the claim for non-pecuniary loss could help account for the failure of PSLA to keep pace with inflation. Even though there had been this cross-over of certain items of loss, the court concluded that there should still be an increase for non-pecuniary loss in serious cases.

The wider economic effects

Although the court was not directly influenced by the substantial rise in damages for pecuniary loss in recent years, it emphasised that it was concerned with the economic consequences of its decision upon not only defendants as a group, but

45 Proposals for the link to be re-established came back to haunt the New Labour government at its party conference last autumn. See "Pensions Haggling Fails To Buy Off Rebels" *The Times*, 28 September 2000

46 [1999] AC 345, [1998] PIQR Q56

also society as a whole. The effect on insurance premiums and upon who would have to pay was a matter of concern. It cited with approval Diplock LJ's view that: 'To avoid fixing the scale at a level which would materially affect the cost of living or disturb the current social pattern is a factor, Benthamite no doubt in origin, in the empirical process by which the maximum/datum is determined.'⁴⁷ Here, therefore, the court differed from the Commission which had attached little importance to these matters and had suggested that cost was irrelevant.

However, the court distinguished its task in assessing PSLA from its assessment of pecuniary loss. In cases such as *Wells* it suggested that the court is only required to make the correct calculation, and economic consequences are irrelevant. But in setting the level of PSLA the court was concerned with wider questions of social policy. In particular, for example, the impact upon the resources of the NHS was regarded as a relevant factor. NHS trusts are among those hardest hit by the *Heil* decision because clinical negligence claims are usually of a high value, being more likely to involve serious injury. For example, in Wales claims over £30,000 were expected to cost £17 million in the current financial year. If the Law Commission recommendations were implemented, a further £5 million would be required, a rise of 30 per cent.⁴⁸ Even taking into account the lower increases set by the Court of Appeal, it is clear that a more onerous burden has now been placed on the NHS. The court considered that it was in a position to appreciate fully the significant effect of its decision on the public at large and the NHS in particular.

Conclusion: the distorted lens of corrective justice

Atiyah's criticisms of PSLA are well known, and require only a brief summary here. PSLA is a secondary form of compensation, incapable of precise assessment and less able than pecuniary damages to provide an equivalent to what has been lost. A 'correction' cannot be made. It is compensation deserving a lower priority than that awarded for financial loss. Few other systems of compensation make provision for PSLA, almost nothing being available from either social security or private insurance. If left to market forces, people would not choose to purchase cover against such a loss. Atiyah's arguments based on distributive justice emphasise how poorly the majority of accident victims are treated because they are unable to claim in tort. He exposes the fault principle as completely inadequate to justify the comparatively generous payments provided to the few whilst excluding the many. A reallocation of resources is required. In order to afford extending the base of compensation there would have to be a limit on the payment of secondary compensation. 'It is perhaps only in the most serious cases of long-term pain and loss of faculty resulting from major physical injuries that there is a good case for damages for non-pecuniary loss.'⁴⁹

47 *Wise v Kaye* [1962] 1 QB 638 at 670. Objection to the relevance of cost was one of the main reasons why an appeal to the Lords was contemplated, as discussed in n 6 above.

48 Outline submission for the respondent in the appeal of *Annable v Southern Derbyshire Health Authority*, one of the *Heil v Rankin* test cases. Actuaries for the NHS Liability Authority calculated that the costs for the NHS in general would rise by £133 million a year, almost as much as the £150 million cost of employing over 8,000 newly qualified nurses. Even without the additional sums for PSLA, damages for clinical negligence rose from £11 million in 1996-97 to £66 million the following year. They were projected to grow to £278 million in 1999-2000. This enormous increase is partly attributable to the revised multiplier resulting from the 3 per cent discount rate set in *Wells v Wells*.

49 P. Cane, *Atiyah's Accidents, Compensation And the Law* 6th ed (London: Butterworths, 1999) 142. Similarly Sugarman, 'Serious Tort Law Reform' (1987) 24 San Diego LR 795, 823-830. Of course,

By contrast the Law Commission rely on its terms of reference to justify giving these distributive justice arguments little space. They do not figure at all in the Court of Appeal judgment. The result is that, almost without argument, corrective justice is accepted as the only lens through which tort law can be viewed. For the unsuspecting student the many deficiencies of tort are masked by a decision which produces 'fair, reasonable and just' compensation supposedly corresponding to the wishes of society. The reliance upon empirical surveys and mathematical indices reinforce the correctness of the approach. Damages are increased and the tort system strengthened. *Heil v Rankin* improves the position of those able to establish liability as against those unable to do so.⁵⁰ It thus helps to entrench the tort system, making it increasingly difficult to revise radically the baseline and extend compensation for basic losses to those presently excluded. In this respect the decision is by no means the most significant case in recent years. However, the case also makes repeated claims to reflect society's sense of values. This article points out what should be obvious enough, that these claims are inevitably partial and depend upon the perspective through which accident compensation is viewed. This partiality requires that the explanation of law reform be in terms of the activity of moral entrepreneurs⁵¹ rather than the general social interest mysteriously given body and voice by both the Law Commission and the court in *Heil v Rankin*.

Atiyah radically changed his views as to how the revision in compensation might be achieved in *The Damages Lottery* (Oxford: Hart Publishing, 1997). The scope for increased use of private insurance, as suggested by Atiyah, is forcefully criticised in an excellent note by Conaghan and Wade in 'From The Permissive To The Dismissive Society' (1998) 25 J Law and Society 284. cf Burrows, above n 38.

50 Cf the reluctance of Lord Bridge in *Hodgson v Trapp* [1989] AC 807, 823 to 'add to the enormous disparity between the position of those who are able to establish a third party's fault as the cause of their injury and the position of those who are not.'

51 See similarly Atiyah's criticism that the Law Commission 'is far too closely wedded to the system and its underlying value structure, to be able to bring to bear the independent scrutiny the system needs.' *The Damages Lottery* n 49 above, 173.

Public Law Representations and Substantive Legitimate Expectations

Melanie Roberts*

Members of the public seek and rely on advice and acts of public bodies and their officials, yet to what extent can public bodies and their officials be held to their representations?¹ The type of problem that can arise is where a public body has made a representation, which is within its powers to make, but it later seeks to resile from it. Or, where a public body has set out policy criteria for dealing with a particular area of decision-making, and an individual has relied on these criteria and legitimately expects a favourable decision based on these criteria, and then the public body changes its mind and seeks to apply different policy criteria to the particular area and so dashes the expectations of those affected. Can the individual obtain the substantive benefit which would be forthcoming if the public body was held to the representation or the original policy criteria was still the public body's chosen option? Is a legitimate expectation aroused in the applicant, and if so, does it give rise only to a procedural protection:² a right to be heard as to whether the original policy criteria or representation should apply? Or does it give rise to substantive protection: the original criteria will be applied or the representation adhered to? This question has generated a considerable amount of debate in the courts³ and amongst commentators.⁴ The recent Court of Appeal decision in *R v North and East Devon Health Authority, ex p. Coughlan*⁵ has brought the doctrine of substantive legitimate expectation back into the spotlight, having had a shadow cast over it by a differently constituted Court of Appeal in *R v Secretary of State for the Home Department, ex p Hargreaves*.⁶

The doctrine of substantive legitimate expectations

The courts have accepted that procedural protection should be given where an individual has a legitimate expectation of procedural protection such as a hearing

* School of Legal Studies, University of Sussex.

1 'Representations' include promises, assertions, acts, practices, policy statements.

2 It is established that the principle of legitimate expectation can give procedural protection, see *Schmidt v Secretary of State for Home Affairs* [1969] 2 Ch 149; *R v Liverpool Corporation, ex p Liverpool Taxi Fleet Operators' Association* [1972] 2 QB 299; *AG of Hong Kong v Ng Yuen Shiu* [1983] 2 AC 629; *Council of Civil Service Unions v Minister for the Civil Service* [1985] AC 374. See C. Forryth, 'The Provenance and Protection of Legitimate Expectations' (1988) 47 CLJ 238; P. Craig, 'Legitimate Expectations: A Conceptual Analysis' (1992) 108 LQR 79.

3 *R v Ministry for Agriculture, Fisheries and Foods, ex p Hamble (Offshore)* [1995] 2 All ER 714, *R v Secretary of State for Transport, ex p Richmond upon Thames BC* [1994] 1 All ER 577, *R v Secretary of State for the Home Department, ex p Hargreaves* [1997] 1 All ER 397, *R v Inland Revenue Commissioners, ex p Unilever plc* [1996] STC 681.

4 n 2 above; Craig, 'Substantive Legitimate Expectations in Domestic and Community Law' [1996] CLJ 289; Singh and Steyn 'Legitimate Expectation in 1996: Where Now?' [1996] JR 17; R. Gordon and T. Ward, 'The Billowing Fog' (1996) NLJ 1663; Foster, 'Legitimate Expectations and Prisoners' Rights: The Right to Get what you are Given', [1997] MLR 727; Bamforth, 'Fairness and Legitimate Expectation in Judicial Review' [1997] CLJ 1; T. R. S. Allan, 'Procedure and Substance in Judicial Review', [1997] CLJ 246; Craig, 'Substantive Legitimate Expectations and the Principles of Judicial Review' in M. Andenas (ed), *English Public Law and the Common Law of Europe* (London: Key Haven Publications plc, 1998).

5 [2000] 2 WLR 622.

6 [1997] 1 All ER 397.

or of a consultation before a decision is made.⁷ Fairness dictates that the expectation of a hearing, or other procedural protection be fulfilled. It is also accepted that where an individual has a legitimate expectation that a benefit of a substantive nature will be granted, or if already in receipt of the benefit, that it will be continued, then fairness too dictates that expectation of the benefit should give the individual the entitlement to be permitted to argue for its fulfilment.⁸ In this situation the decision-maker merely has to hear what the individual has to say but does not have to give the substantive benefit. What has been the subject of some controversy is whether or not a legitimate expectation that a benefit of a substantive nature will be granted, or if already in receipt of the benefit, that it will be continued, can give rise to substantive protection, in other words to the individual obtaining or continuing to receive the benefit. This is what is referred to as substantive legitimate expectations.

Laws J in *R v Secretary of State for Transport, ex p Richmond upon Thames London BC*⁹ exemplifies the traditional objection to the doctrine of legitimate expectations. He held there could be no doctrine of substantive legitimate expectation because it 'would impose an unacceptable fetter upon the power of a public body to change its policy when it considered that that was in fulfilment of its public responsibilities.'¹⁰ The argument against the doctrine of substantive legitimate expectation is that public bodies with discretionary powers can not be bound by prior representation or policy choice, for if they are, they fetter their discretion. Such a fetter on discretion would itself be ultra vires. If a body is held to an ultra vires representation it would threaten the whole ultra vires doctrine: allowing public bodies to extend their powers at will. Therefore once a public body seeks to resile from an earlier representation or changes its policy criteria it can no longer be held to the representation or earlier policy option. Although neither policy choice is ultra vires in the sense of being outside the power of the public body, the ultra vires principle manifests itself because if a public body fetters its discretion, it will have acted beyond its powers. The principle of legality, as manifested in the ultra vires doctrine, dictates that an applicant will fail to obtain substantive protection because discretion can not be fettered. Public bodies must be free to change and develop their policies.

However, there is another value which must be acknowledged: the principle of legal certainty. This principle is recognised in European Community law.¹¹ The doctrine of legitimate expectation is based upon the principle of legal certainty. This principle dictates that if a public body represents that it will behave in a particular way, then the individual expects that representation to be complied with. In other words, where a promise is made by a public body it is in the interests of good administration that it should act fairly and should implement its promise. The idea that legitimate expectations ought not to be thwarted is founded upon a basic principle of fairness.¹² There is an increasing awareness of the principle of substantive fairness in public administration. Sedley J in *R v Ministry for*

7 *R v Liverpool Corporation, ex p Liverpool Tax Fleet Operators' Association*, n 2 above; *AG of Hong Kong v Ng Yuen Shiu*, n 2 above.

8 *R v Secretary of State for Transport, ex p Richmond upon Thames London BC*, n 3 above.

9 n 3 above.

10 n 8 above, 596.

11 Craig, 'Substantive Legitimate Expectations in Domestic and Community Law', n 4 above and Craig, 'Substantive Legitimate Expectations and the Principles of Judicial Review', n 4 above.

12 De Smith, Woolf and Jowell, *Judicial Review of Administrative Action*, (London: Sweet & Maxwell, 1995) 417, para 8.038.

*Agriculture, Fisheries and Foods, ex p Hamble (Offshore) Fisheries Limited*¹³ bases the doctrine of substantive legitimate expectation on fairness in the public administration, 'it is difficult to see why it is any less unfair to frustrate a legitimate expectation that something will or will not be done by the decision-maker than it is to frustrate a legitimate expectation that the applicant will be listened to before the decision-maker decides whether to take a particular step.'¹⁴ It is in the interests of good administration that a body should act fairly.¹⁵

The principles of legality and legal certainty conflict with one another,¹⁶ the former seeks to deny any doctrine of substantive legitimate expectation, whilst the latter expects substantive protection for the applicant where an expectation has been aroused. Clearly the administration should not be fettered in its duty in acting in the best interests of the public, but equally the individual should be protected where it has relied on a representation of the administration. In order for the doctrine of substantive legitimate expectations to be recognised there must be acknowledgement of these conflicting interests, and the existence and importance of the value of legal certainty must be accepted.

In the common law of the European Union it is accepted that expectations can give rise to substantive protection, with no distinction being made between procedure and substance. The European Court of Justice (ECJ) recognises the doctrine of substantive legitimate expectation; it acknowledges that there are conflicting interests at stake, that of the individual who has relied on the earlier policy, and that of the public body which must be free to change its policy.¹⁷ Once the ECJ is satisfied that a legitimate expectation exists, which has arisen either from a representation or conduct of the public body, the needs of the individual are balanced with those of the public body, so that provided there is no outweighing harm to the public interest, the legitimate expectation of a substantive benefit will be upheld.

Status of doctrine of substantive legitimate expectations prior to *ex parte Coughlan*

Prior to the Court of Appeal's decision in *ex p Hargreaves*, the weight of authority was in favour of the developing doctrine of substantive legitimate expectations. Indirect authority was given for the recognition of a legitimate expectation giving rise to substantive protection in *R v Secretary of State for the Home Department, ex p Khan*.¹⁸ Parker LJ held that the Secretary of State should not be allowed to renege from stipulated conditions of entry 'without affording interested persons a hearing *and* then only if the overriding public interest demands it'¹⁹ (emphasis added). There is clear recognition of both a procedural aspect, a need for a hearing, and a substantive aspect, the requirement that the public interest demands that the

13 n 3 above.

14 *ibid*, 724.

15 See Lord Mustill in *R v Secretary of State for the Home Department, ex p Doody* [1993] 3 All ER 92, 106.

16 See J. Schwarze, *European Administrative Law* (London: Sweet & Maxwell, 1992) 867–868.

17 For discussion of EC law see n 16 above, 1134–1135 and see Sedley J in *R v Ministry for Agriculture, Fisheries and Foods, ex p Hamble (Offshore)* [1995] 2 All ER 714; Craig, 'Substantive Legitimate Expectations in Domestic and Community Law', n 4 above, 304–310, Craig, 'Substantive Legitimate Expectations and the Principles of Judicial Review', n 4 above.

18 [1984] 1 WLR 1337.

19 *ibid*, 1344.

pre-existing policy must be departed from in relation to the individual applicant.²⁰ *R v Secretary of State for the Home Department, ex p Ruddock*²¹ was precisely a case of a substantive legitimate expectation.²² The applicant claimed she had a legitimate expectation that published criteria would be followed when a decision was taken as to whether to intercept her telephone calls. This is clearly a case with no procedural dimension; there could be no hearing before an inception order is made (it would frustrate the purpose of such an order if notice was given of an impending interception).

Sedley J's (as he then was) judgment in *ex p Hamble* was a clear endorsement of the doctrine of substantive legitimate expectations. He held that the balance between the individual's expectation and the public interest is in the first place for the policy maker, but if the outcome is challenged by judicial review then it is for the courts to decide whether there is an overriding reason for frustrating the legitimate expectation, 'it is the court's task to recognise the constitutional importance of ministerial freedom to formulate and to reformulate policy; but it is equally the court's duty to protect the interests of those individuals whose expectation of a different treatment has a legitimacy which in fairness outtops the policy choice which threatens to frustrate it.'²³

However, the court in *ex p Hargreaves* relied entirely on *Findlay v Secretary of State for the Home Department*,²⁴ a case which also involved prisoners, in order to reject the applicants' claim that they had a legitimate expectation based on the earlier policy. The court cast a shadow over the doctrine of substantive legitimate expectation by suggesting that it was not for the court to determine the fairness of a minister's decision not to accommodate a reasonable expectation which a policy would thwart, as this amounted to an intrusion into the merits of the decision. It was suggested that on matters of substance *Wednesbury* is the correct test, and that the doctrine of legitimate expectation, based on fairness, cannot be extended from procedural to substantive matters.²⁵ It was for the decision-maker to undertake the balancing act: to decide whether the expectation should be protected or whether the public interest is strong enough to override the expectation. The court would only quash the decision to apply the new policy instead of the old, if it could be shown that the decision maker's judgment to do was irrational, *Wednesbury* unreasonable; it could not be quashed on the basis of fairness. The court in *Hargreaves* overrode Sedley J in *Hamble* in so far as he propounded that a balancing exercise should be undertaken by the court, going so far as to declare it heretical.²⁶

It was unclear what subsequent courts would do following *Hargreaves*. Would *Hargreaves* be interpreted as authority against the principle of legitimate expectations giving rise to substantive protection, so that it would not be possible for any one to argue for substantive protection? Or would the case be limited to its facts?

20 see Craig, 'Substantive Legitimate Expectations in Domestic and Community Law', n 4 above, 296 and see Sedley J in *R v Ministry for Agriculture, Fisheries and Foods, ex p Hamble (Offshore)*, n 3 above, 731.

21 [1987] 1 WLR 1482.

22 n 13 above, 723.

23 *ibid*, 731.

24 [1985] AC 318.

25 n 6 above, 412 and 416.

26 *ibid*, 415.

***Ex parte Coughlan* and the doctrine of substantive legitimate expectation**

A differently constituted court of appeal has now had the opportunity of deciding the status of the doctrine of substantive legitimate expectations in English law in *R v North and East Devon Health Authority, ex p Coughlan*. The judgment in *Coughlan* was contributed to by all three judges who heard the case, Lord Woolf and Lord Justices Mummery and Sedley and was handed down by Lord Woolf. Both Lord Woolf and Sedley LJ support for the doctrine of substantive legitimate expectation was well known prior to this case.²⁷

The facts of *Coughlan* were these: in 1993 Miss Coughlan, who was severely disabled, was moved, along with seven others, from Newport hospital, which the Health Authority wished to close, to Mardon House, a National Health Service facility for the long-term disabled. At the time of the move the Health Authority gave an express assurance, both verbally and in writing, that Mardon House would be Miss Coughlan's and her fellow patients' home for life. In 1998, following public consultation, the Health Authority decided to close Mardon House and to transfer the long-term general nursing care of Miss Coughlan to the local authority, without any alternative placement for her being identified. The case gave rise to a number of issues, including the question of what was the effect in law of the Health Authority's promise that Mardon House would be Miss Coughlan's home for life. Could the courts uphold a legitimate expectation of a substantive benefit?

In a clear and well reasoned judgment, the Court of Appeal started by acknowledging that the court's role is still the subject of some controversy where an individual has a legitimate expectation of being treated in a particular way by a public body and the public body wishes to treat him or her in a different way. The Court of Appeal stated that the starting point is to ask what the individual's *legitimate* expectation was, and suggested that where there is a dispute as to this it is to be determined by the court, with there being at least three possible outcomes with the court taking a different role in respect of each category. In the first category the court may decide the public body only needs to bear in mind its previous policy or assurances, giving it the weight the public body thinks right when deciding whether to change course. The court will then only review the decision on conventional *Wednesbury* grounds. *Hargreaves* is an example of a case falling within this category. The second category is where the court decides the representation gives rise to a legitimate expectation of a procedural benefit eg consultation, then the court will require the opportunity for consultation to be given unless there is an overriding reason to resile from it. The court here is determining, again on conventional grounds, whether the decision was procedurally fair.

The third category is where a legitimate expectation has been aroused of a benefit which is substantive. The court will, in a 'proper case', decide whether to frustrate the expectation is so unfair that to take a new and different course will amount to an abuse of power.²⁸ Once the legitimacy of the expectation is established, it is for the court to determine whether there is a sufficient overriding interest relied upon for the change of policy or to justify departing from the promise. The court is undertaking a balancing exercise between the public interest

27 Lord Woolf, n 12 above, shows support for the doctrine of substantive legitimate expectation and Sedley J in *Hamble* had made clear his support for the doctrine.

28 n 5 above, 645.

and the individual's interest. This third category²⁹ is clear acceptance of the doctrine of substantive legitimate expectation: a legitimate expectation of a substantive benefit giving rise to protection of the benefit, not simply procedural protection. The court defines the type of case of an enforceable expectation of a substantive benefit as being where the expectation is confined to one person or a few people, giving the promise or representation the character of a contract.³⁰ The present case was held to come within this category as the promise given by the health authority was limited to a few people. What was also held to be important, was that the consequences to the health authority of being held to its promise were likely to be financial only.³¹

Promises, rather than policies, are more likely to fall within the third category. It is easier to uphold a legitimate expectation of a substantive benefit in respect of a promise than in respect of a policy. As Craig has pointed out, within the EC legal system and in many continental systems, the type of substantive legitimate expectation case that is treated as the strongest, is where an individualised representation has been made and then the administrative body changes its mind and makes an individualised decision which is inconsistent with the original representation.³² This type of case is treated as the strongest: 'because an unequivocal representation made to a person carries a particular moral force and because holding the public body to that representation is less likely to have serious consequences for the administration as a whole.'³³ *Coughlan* is precisely this type of case.

The main difference between promises and policies are that promises are specifically directed to one or a few individuals stating what the public body will do in their particular case. As Sedley J said in *Hamble*, 'a promise is precisely, a representation about future conduct, making it relatively straightforward to decide whether the promisor should be held to it. Practices may, but do not necessarily, have the same character. Where the material practice is generated by a policy which is itself liable to change, the practice cannot be logically expected to survive a policy change...'³⁴ The courts are going to be more prepared to intervene in cases where the applicant was given a promise. *Coughlan* then is the ideal case to uphold a substantive legitimate expectation claim as it involved a promise to only a few individuals. Where policies are involved it is harder for the courts to sustain a substantive legitimate expectation claim as the courts must be careful that upholding such a claim based on an original policy does not lead to the fettering of policy.

The difficulty which may arise however is in the categorisation of cases; the difference between a policy and an individual promise may not always be clear and litigation could ensue over definitions.³⁵ The court in *Coughlan* is not suggesting that all policy type cases will come within the first category. The court acknowledges that deciding into which category a decision should be allotted

29 A fourth outcome would be where there is a legitimate expectation of a substantive benefit which gives rise to procedural protection, usually a hearing as to whether or not the original policy should be applied to the individual who relies on an expectation of it applying to them.

30 n 5 above, 646.

31 *ibid*.

32 Craig, 'Substantive Legitimate Expectations and the Principles of Judicial Review', n 4 above.

33 *ibid*, 26.

34 *R v Ministry for Agriculture, Fisheries and Foods, ex p Hamble (Offshore)*, n 3 above, 729, and see Yoav Dotan, 'Why Administrators should be Bound by their Policies' OJLS 17 [1997] 23, 41; Craig, 'Substantive Legitimate Expectations in Domestic and Community Law', n 4 above, 294.

35 n 32 above, 27.

will be a difficult task.³⁶ Subject matter will necessarily play a part. *Findlay* and *Hargreaves* both involved prisoners and the policies affected more than 'a few individuals'. It would, of course, have been possible in *Findlay* and in *Hargreaves* for the court to hold that a legitimate expectation had been aroused in the applicants but that in both these cases it could not be upheld because there was an overriding public interest: the importance of the public interest in the administration of parole and home leave.

Acknowledging that the third category is still controversial the court relies upon two taxation cases, the House of Lords decision in *Preston v Inland Revenue Commissioners*³⁷ and the Court of Appeal decision in *R v Inland Revenue Commissioners, ex p Unilever plc*.³⁸ In both these cases the question which had been asked by the courts was whether frustrating a legitimate expectation can amount to an abuse of power. Whilst the language of substantive legitimate expectations was not used in these two cases, it was nonetheless clear that in talking about 'an abuse of power' the courts were giving tacit recognition to the doctrine of substantive legitimate expectation. Until *Coughlan*, the vocabulary of 'abuse of power' had not been used in fields of public law other than those involving the Inland Revenue. *Preston* was not even referred to in *Hargreaves*. In *Preston* whilst the applicants claim, that the Inland Revenue should honour an agreement not to pursue tax claims, failed on the facts, Lord Templeman acknowledged that in principle a breach of conduct or a breach of representation on the part of a public body is unfair and could amount to an abuse of power.³⁹ In another case involving the Inland Revenue, *R v Inland Revenue Commissioners, ex p Matrix Securities Ltd*,⁴⁰ the House of Lords accepted that in appropriate circumstances a legitimate expectation can give rise to substantive protection, that it may be an abuse of power for the Inland Revenue to seek to extract tax contrary to an advance clearing given by the Inland Revenue.⁴¹

In *ex p Unilever plc*, Unilever argued that the Inland Revenue should have allowed them to claim tax relief on trading losses incurred within a particular year. The Inland Revenue refused to allow Unilever's claim on the basis that it was outside the two year time limit during which losses should be claimed. Unilever asserted that as the Inland Revenue had, over the previous twenty years, allowed them to claim the relief outside the two year period, they had a substantive legitimate expectation that it would continue to follow its previous practice of allowing late claims. The Court of Appeal upheld Unilever's application. The judgments talk about 'unfairness' and 'abuse of power' in allowing the Inland Revenue to resile from an administrative practice,⁴² using the language of earlier cases which support the protection of substantive legitimate expectation. Unfairness was what made the Inland Revenue's decision unlawful. So that despite concluding that the decision of the Revenue was so unreasonable as to be irrational, the court appears to be saying that it was 'unfair' and 'an abuse of power' *because* Unilever had a legitimate expectation derived from past practice of

36 n 5 above, 645.

37 [1985] AC 835.

38 [1996] STC 681.

39 n 37 above, 866-867.

40 [1994] 1 WLR 334.

41 See comments of Lord Woolf in *New Zealand Maori Council v Attorney General of New Zealand* [1994] 1 AC 466 regarding recognition of substantive legitimate expectations.

42 n 38 above and the judgments of Sir Thomas Bingham MR and Simon Brown LJ.

not insisting upon the strict time limit. It is the breach of the legitimate expectation which is an abuse of power.⁴³

The court in *Coughlan* recognises the delicate balance that must be achieved between not preventing public bodies from remaining free to change their policies, but also needing to protect the public when their legitimate expectations, which were aroused by the public body, are being frustrated. Whilst the court does not itself discuss explicitly the principles of legality and legal certainty, it notes that De Smith, Craig and Schwarze link substantive legitimate expectations to the fundamental principle of legal certainty.⁴⁴ It is the principle of legality that lies behind the need to ensure public bodies are free to change their mind in the public interest, and the principle of legal certainty that lies behind the need to ensure that individuals are protected when they rely upon a representation of a public body. The critical question is, by what standard the court is to resolve such conflicts.⁴⁵ *Coughlan* stated that whilst the court does not usually intervene where a decision has passed the hurdles of rationality and proper purpose, it acknowledged that this type of case is different. *Coughlan* involved two lawful exercises of power: the promise and the policy change, with individuals being trapped between the two.⁴⁶ The role for the court is therefore 'to ensure that the power to make and alter policy has not been abused by unfairly frustrating legitimate individual expectations. In such a situation a bare rationality test would constitute the public authority judge in its own cause, for a decision to prioritise a policy change over legitimate expectations will almost always be rational from where the authority stands, even if objectively it is arbitrary or unfair.'⁴⁷ The court suggests that the approach, following *Preston*, is to ask whether the unfairness in frustrating the legitimate expectation amounts to an abuse of power. Fairness includes fairness of outcome,⁴⁸ in other words, substantive fairness. *Preston* is interpreted as making clear that 'abuses of power include reneging without adequate justification, by an otherwise lawful decision, on a lawful promise or practice adopted towards a limited number of individuals.'⁴⁹

Support for the assertion that unless there is an overriding public interest the legitimate expectation of a substantive benefit should be upheld, can be found in Lord Denning's judgment in *HTV Ltd v Price Commission*, 'it is a misuse of power for (a public body) to act unfairly or unjustly towards the private citizen when there is no overriding public interest to warrant it.'⁵⁰ A balancing exercise needs to be carried out by the courts. The court in *Coughlan* looks at a number of authorities and makes the point that nowhere is 'there any suggestion that judicial review of a decision which frustrates a substantive legitimate expectation is confined to the rationality of the decision.'⁵¹ What was said in *Hargreaves* in respect of the proper approach for the court to take in reviewing a decision of a public body, was strictly obiter, and the court noted had been strongly questioned by a number of commentators.⁵² *Hargreaves* was, in any event, distinguished from *Coughlan*, as in *Coughlan*, 'fairness in the statutory context required more of the decision-maker

43 See R. Gordon and T. Ward, 'The Billowing Fog', n 4 above.

44 n 5 above, 653.

45 *ibid*, 647.

46 *ibid*, 648.

47 *ibid*.

48 *ibid*, 649.

49 *ibid*, 648.

50 [1976] ICR 170, 185.

51 n 5 above, 651.

52 *ibid*, 652.

than in *Hargreaves*'.⁵³ *Hargreaves* can be confined to the facts of its case. *Coughlan* made clear that whether there is an overriding public interest is a question for the court.⁵⁴ Whilst lawful policy change is always for the public body, the courts' task is 'limited to asking whether the application of the policy to an individual who has been led to expect something different is a just exercise of power'.⁵⁵ It is for the court to decide whether frustration of the individuals expectation is so unfair as to be a misuse of the authorities power.⁵⁶

In *Coughlan* the court found that an express promise had been made to Miss Coughlan that Mardon House would be her home for life. Miss Coughlan had relied on this promise. The court held that strong reasons were required to justify resiling from a promise in these circumstances.⁵⁷ The reason that the health authority gave for wanting to close Mardon House was to save money and financial consequences were not viewed by the court as an overriding public interest.⁵⁸ It is apparent that in *Coughlan* the promise that was made to Miss Coughlan was seen as akin to a contract and the health authority had not offered her an equivalent replacement facility. The decision to move Miss Coughlan was held to be unfair because it frustrated her legitimate expectation of having a home for life in Mardon House and there was no overriding public interest which justified it. The unfairness therefore amounted to an abuse of power by the health authority.

Using the language of 'abuse of power' and 'fairness', the Court of Appeal has brought the doctrine of substantive legitimate expectations, within UK domestic law, more into line with that in the European Union. Where a public body has made a representation (usually a promise), that body can be held to the representation where it is substantively unfair, amounting to an abuse of power, to frustrate the legitimate expectation that has been aroused by the representation and there is no overriding public interest to justify resiling from the representation.

Ultra vires representations

The cases discussed above all concerned intra vires representations, in other words the promises, practices and policies were all within the statutory power of the public body. But what of situations where a public body makes a representation which the body has no power to make, or whilst the public body has power to make the representation the officer who makes it does not have the power to do so? Such a representation can generate an expectation in an individual and yet an individual who relies on such a representation, will not, in most cases, have any reason to suspect or know that the representation was ultra vires either the officer or the public body, however traditionally the courts have denied the individual any relief.

There are a number of reasons which have been given as to why an ultra vires representation should not be binding on the body which made it. The most oft-repeated rationale is that stated by Lord Greene, 'the power given to an authority under a statute is limited to the four corners of the power given. It would entirely destroy the whole doctrine of ultra vires if it was possible for the donee of a statutory

53 *ibid*, 652.

54 *ibid*.

55 *ibid*, 654.

56 *ibid*.

57 *ibid*, 656.

58 In many cases involving resource allocation the court has refused to interfere with the decision, eg *R v Cambridge Health Authority ex p B* [1995] 2 All ER 129.

power to extend his power by creating an estoppel.'⁵⁹ In other words, if the courts had power to dispense with those limits in order to give relief to an individual who has relied on a representation which is outside the power of the body or the officer who made it, this would make a nonsense of the idea of power limited by statutory provision. Related to this reason is the idea of separation of powers. The doctrine of ultra vires defines the boundaries between the powers of the legislature, the courts and the administration. If the courts were to dispense with the ultra vires doctrine this would entail a shift of power to the courts away from the legislature and the administration; the courts would be usurping Parliamentary Sovereignty.

The principle of legality is cited to prevent public bodies being held to unlawful representations upon which an individual has relied. The purpose of the ultra vires doctrine is to protect the public from public bodies unlawfully extending their powers. In this sense the ultra vires concept is the embodiment of the principle of legality. However what of the principle of legal certainty? Where an innocent individual has relied upon an ultra vires representation, not knowing that the body or officer did not have the power to make the representation, should the individual have no redress? Allowing the principle of legality to automatically override that of certainty could cause the individual to suffer severe detriment.

The doctrine of estoppel was used by Denning J (as he then was) in *Robertson v Minister of Pensions*⁶⁰ to give relief to an individual who had relied on an unlawful representation. Robertson, a disabled army officer relied, to his detriment, on a War Office assurance that his disability was attributable to war service and did nothing about a private pension. Subsequently the Minister of Pensions refused to accept the disability was attributable to war service. However Denning J held that the Minister of Pensions was bound by the War Office and was therefore estopped from resiling from that assurance. The powers of the Minister of Pensions were not exercisable by the War Office and therefore the War Office assurance could have been regarded as ultra vires and therefore a nullity. However here the court is not concerned with Parliamentary Sovereignty and the jurisdictional principle, but with the legal certainty argument, 'if a government department in its dealings with a subject takes it on itself to assume authority on a matter with which (the applicant) is concerned, he is entitled to rely on it having the authority which it assumes. (The applicant) cannot be expected to know the limits of its authority'.⁶¹ However the House of Lords in *Howell v Falmouth Boat Construction Co*⁶² disapproved of Denning LJ's remarks relating to an ultra vires assurance and its legal consequences,⁶³ holding that 'the illegality of an act is the same whether or not the actor has been misled by an assumption of authority on the part of a government officer however high or low in the hierarchy'.⁶⁴ So that if an assurance is ultra vires the authority it does not bind the authority.⁶⁵

If it is discovered in time that a mistake has been made and an assurance has been given which is ultra vires the body or officer who made it, then it can be corrected

59 *Minister of Agriculture and Fisheries v Hulin*, unreported but cited in *Minister of Agriculture and Fisheries v Matthews* [1950] 1 KB 148, 154.

60 [1949] 1 KB 227.

61 *ibid*, 232.

62 [1951] AC 837.

63 n 69 above and in the Court of Appeal in *Howell v Falmouth Boat Construction Co* [1950] 2 KB 16, where Denning held that an applicant should be entitled to rely on acts of public authorities and ought not to suffer if they exceed their authority.

64 *ibid*, 845 per Lord Simonds.

65 See *Western Fish Products Ltd v Penwith District Council* [1981] 2 All ER 204 for a discussion of estoppel in public law.

before it has been acted upon.⁶⁶ However if the applicant has reasonably relied upon an ultra vires assurance should the public body be held to the representation or should legality always override legal certainty and the applicant suffer loss?⁶⁷ One solution would be to give compensation to the representee rather than allow an ultra vires representation to bind a public body. Support for this comes from Lord Griffiths in *Matrix Securities* where, obiter, he suggested that had the applicant been entitled to rely upon the assurance given by the local inspector of taxes 'fairness demands that the applicant should be reimbursed for this out-of-pocket expense and it could be regarded as an abuse for the Revenue to refuse to do so.'⁶⁸ Paying compensation ensures that legality is preserved but what of the principle of legal certainty? It will not always be the case that the harm to the public, who are the beneficiaries of the ultra vires doctrine in allowing the representation to stand, will outweigh the harsh effect to the individual in disallowing the representation to stand. For example the representation in *Robertson* does not affect third parties, and the individual harm to Robertson of not having his expectations fulfilled outweighs any harm to the public in having tax revenue spent on Robertson's pension. The principle of legal certainty outweighed that of legality. There are echoes here of the balancing act that is undertaken with substantive legitimate expectations. Where the representation relates to a tax or social security matter the harm to the individual will usually outweigh any harm to the public.

The reasoning in *Coughlan* could be applied in order to give relief in cases where an individual has relied upon a promise made by a public body and the individual has no reason to know the representation was ultra vires. In deciding whether to allow the public body to resile from the promise the court would need to balance the public interest and the private interest; legality and certainty. If the private interest outweighs the public interest then the court should hold that it would be an abuse of power for the public body to resile from the representation.

Conclusion

The decision of the Court of Appeal in *Coughlan* and the recognition that there can be substantive legitimate expectations is to be welcomed. Whilst *Coughlan* categorises legitimate expectation cases into three different types, all cases involving representations made by public bodies have the same underlying problem: conflict between the public interest and the individual interest; between the interests of legality and legal certainty. What matters in the common law of the EC and of Britain is whether the applicant can demonstrate 'an expectation which is worthy of protection'.⁶⁹ It will be easier to do this where a public body has made a promise. However *Coughlan* is not suggesting that an expectation aroused by a policy change cannot be upheld. It will depend on whether there is a sufficient overriding public interest to justify departing from the original policy in respect of the applicant. Whilst *Coughlan*, *Preston* and *MFK* all involved express promises made by a public body which then changed its mind and made a decision which was inconsistent with the original representation, *Unilever* also allowed relief even though the representation involved a practice which had continued for some time.

66 See *R v Inland Revenue Commissioners, ex p Matrix Securities Ltd.* [1994] 1 All ER 769, 781 per Lord Griffiths.

67 See Craig, 'Substantive Legitimate Expectations in Domestic and Community Law' n 4 above, 311 for a discussion of the position with EC law.

68 n 66 above, 782.

69 J. Schwarze, *European Administrative Law* n 6 above at 1134 quoted by Sedley J in *Hamble*, n 27 above, 728.

REVIEW ARTICLE

The Developing International Law of Democracy

*Richard Burchill**

Gregory H. Fox and Brad R. Roth (eds), Democratic Governance and International Law. Cambridge: Cambridge University Press, 2000, 585 pp, pb £19.95.

I

Since the early 1990s democracy has become a topic of great interest in international law and relations. The end of communism in the Soviet Union and elsewhere has been described as the international victory of democracy.¹ This 'victory' of democracy across the world quickly led to claims that there now exists a right to democracy in international human rights law, as well as the existence of democracy as a guiding principle of general international law.²

These developments are welcomed as of positive benefit to the international system in the long-term. However, supporters of democracy as an international legal principle have shown a tendency towards an unquestionable acceptance of the existence of new legal rules that appeared almost overnight. This sudden appearance, and unquestioning acceptance, creates numerous difficulties. The most obvious is the lack of investigation into or discussion of what this idea of democracy contains, and the basis upon which any action in its name may be claimed. Before the collapse of the communist regimes in Europe most states in the world claimed to be democratic in one way or another, even though there were drastic and noticeable differences between the various forms of democracy. Owing to the variety of democratic practices and the fact that 'one word passed as common currency in several different languages', democracy was once an idea not taken seriously at the international level except in the context of ideological struggles.³ Now, it is claimed, the situation has changed and everyone is speaking the same democratic language allowing for a single notion of democracy to be seen as part of international law and relations. It is undeniable that there has been a shift towards democracy in the international system, yet what it means to be democratic or how democracy is manifested in law and practice are issues not adequately

* School of Law, University of Hull. Thanks to Richard Barnes and Rob Cryer for comments on an earlier draft. Numbers which appear in parenthesis in the text are references to the page numbers of the volume under review.

- 1 See Francis Fukuyama, *The End of History and the Last Man* (London: Hamilton, 1992).
- 2 Thomas Franck, 'The Emerging Right to Democratic Governance' (1992) 86 *American Journal of International Law* 46; Anne-Marie Slaughter, 'International Law in a World of Liberal States' (1995) 6 *European Journal of International Law* 503.
- 3 See Susan Marks, 'The End of History? Reflections on Some International Legal Theses' (1997) 8 *European Journal of International Law* 449.

investigated by international lawyers.⁴ The few investigations that have been undertaken tend to be satisfied with the existence of elections as signifying the existence of democracy. This simplistic view of democracy fails to fully embrace the impact and potential democracy has for individuals and societies. The difference between unquestioned assumptions of the victory of democracy in the world and the real and actual events in the day-to-day lives of many individuals often means that 'the New World Order appears more like a boot in the face than a triumph of democracy and the rule of law.'⁵

A second difficulty that arises is the almost instantaneous creation and acceptance of international legal rules that embrace democracy. The emergence of this difficulty has resulted in a general lack of critical perspective concerning the creation and development of international legal principles of democracy.⁶ Apparently the content and impact of democracy as an international legal principle is self-evident. Worrying in this regard is the fact that claims for international legal principles of democracy are based on substantive treaty provisions and state practice, which instantly raises the question as to why discussion of such principles was not occurring sooner. No one laid claim to a right to democracy in 1980, but, based on the evidence provided by many, its status as a right can be dated to 1948.⁷ What has resulted is a retroactive emergence of a legal principle of democracy. It is significant that state practice has only recently allowed for discussions of democracy in international law. However, we must acknowledge that even though the basis for democracy as a principle in international law may have an established history, it has never been embraced to the extent it is today.

II

The recent global process of democratisation has fuelled the growth of democracy as a topic of international law.⁸ The current trend embracing democracy is differentiated from past developments in the extent to which there are 'political changes moving in a democratic direction' at all levels of society – locally, nationally, regionally and globally.⁹ The impact of democracy in the international system is widespread, as international institutions are becoming more active in requiring that their member states maintain democratic systems.¹⁰ At the local level, individuals and groups are calling for greater participation in the processes that impact on their lives.¹¹ Concerns about good governance based on democracy

4 David Held, *Models of Democracy*, 2nd ed (Oxford: Polity, 1996) 291.

5 Derek Spitz, 'Solidarity and Disaggregated Property Rights: Roberto Unger's Recasting of Democracy' (1991/2) 23 *Columbia Human Rights Law Review* 43, 47.

6 Notable exceptions are included in Part V of the volume under review: see below.

7 For example see Christina Cerna, 'Universal Democracy: An International Legal Right or the Pipe Dream of the West?' (1995) 27 *New York University Journal of International Law and Politics* 289, and W. Michael Reisman, 'Sovereignty and Human Rights in Contemporary International Law' (1990) 84 *American Journal of International Law* 866.

8 See 'Democratization and International Law. Building the Institutions of Civil Society' (1994) 88 *ASIL Proceedings* 197.

9 David Potter, 'Explaining Democratization' in Potter et al (eds), *Democratization*. (Cambridge: Polity, 1996) 3.

10 Membership requirements for democracy exist within the Council of Europe, Art 3 and 8 of the Charter; the European Union, Art 6 and 7 of the Treaty of European Union, and the Organisation of American States, Art 9 of the Charter.

11 David Held, 'Cosmopolitan Democracy and the Global Order: Reflections on the 200th Anniversary of Kant's "Perpetual Peace"' (1995) 20 *Alternatives* 427.

are taking a prominent place in the global economy, international development, issues of international trade, aid packages for developing societies and in the general conduct of inter-state relations.¹² Furthermore, non-governmental organisations (NGOs) have proliferated in the new resurgence of democracy playing a major role in the development of democracy at the international level. NGOs are able to act as forums for debate on a wide variety of issues and as pressure groups beyond the constraints of the state structure. Their presence and proliferation at a global level gives the impression of the development of an international civic society.

The widespread development of international legal principles of democracy is evidenced by the twenty essays that make up the present volume. Much of the work contained in *Democratic Governance and International Law* has been published elsewhere but the authors have, for the most part, updated or modified their pieces. The essays collected here represent the vanguard of discussions dealing with democracy as a legal principle. Some offer unwavering support for democracy, others are sceptical of the full range of claims being made in its name. The essays contained in the volume represent the foundation for the future development of and debate about democracy as an international legal principle. They show where problems or inconsistencies lie or where more work needs to be done.

The greatest weakness of the majority of approaches to democracy as an international legal principle is the belief in a limited view of democracy as elections. The belief at the international level that elections alone are satisfactory for the existence of democracy follows the same inadequate approach of democratic thinkers at the domestic level who see democracy as only a procedure. The favouring of a limited procedural view of democracy is attributable to the work of Joseph Schumpeter and has had a considerable influence on the bulk of modern democratic theory and practice.¹³ A procedural view of what democracy is or should be is perhaps the easiest approach to the subject as it provides a verifiable means of determining state practice. However, the desire of international law for the easiest way does not lead to the development of an international law of democracy that is effective or acceptable. Identifying specific practices as being democratic does not indicate the existence or even acceptance of an international legal principle of democracy. Within the international system there is a good deal of rhetoric in favour of democracy. This rhetoric is the product of the increased emphasis being placed on the existence of democracy along with human rights in the international system with an emphasis upon the issue of legitimacy of governments and institutions.¹⁴ In response to this rhetoric there has been an increase in electoral activity throughout the world. However elections alone do not provide for effective democracy, something many of the contributors to the present volume overlook. We cannot be satisfied with a procedural form of democracy without engaging in investigations and debate as to what the procedures mean or provide. The impact and significance of democracy extends beyond elections and

12 For the problems related to attaching democracy to good governance see Katarina Tomaševski, *Between Sanctions and Elections: Aid Donors and their Human Rights Performance* (London: Pinter, 1997).

13 For the influence of Schumpeter on democratic theory see David Ricci, 'Democracy Attenuated: Schumpeter, the Process Theory, and American Democratic Thought' (1970) 32 *Journal of Politics* 239–240; C.B. Macpherson, *The Life and Times of Liberal Democracy* (Oxford: Oxford University Press, 1977) 77–78.

14 Tomaševski, 12 above, 7–14.

the political sphere of society. To ignore the larger impact democracy has on our lives will only lead to disillusionment and disempowerment, the inverse of democracy.

Discussions of democracy as an international legal principle cover a wide variety of international legal topics. Democracy has been conceptualised as an entitlement all societies possess, as a human right individuals are able to exercise through accepted procedures, as a criterion for the recognition of legitimate governance, as a justification for intervention and the use of force, and as an overriding principle upon which the international system is ordered. What is clear is that democracy does not exist as a discrete idea in international law. It has become part of a wide range of activities involving individuals, states and international organisations. Therefore the essays in this volume do not address a single principle of democracy. The various roles democracy has in the international system enable us to speak of the development of an 'international law of democracy'. This international law of democracy contains a variety of principles and ideas that have the purpose of regulating distinct areas of action. As the essays in this volume demonstrate the development of an international law of democracy is undoubtedly in its formative stage.

III

Underlying these various perspectives on democracy is the implicit belief that democracy will provide a path to peace in the international system. Owen provides an in-depth treatment of the democratic peace proposition, a topic that is heavily debated.¹⁵ The idea of a democratic peace in the international system is not new, as its foundations may be traced to Kant's work, *The Perpetual Peace*. Supporters of the notion of democratic peace normally fail to take account of the limits of democracy in creating a more peaceful world. The democratic peace idea holds that democratic states will not go to war with one another. It fails to account for the fact they have no difficulty in going to war with states they deem to be undemocratic. Owen concedes this point when he observes that democracies promote good relations with other democracies but confrontation with states considered undemocratic (p 344). The question then becomes: should we engage in violent confrontational relationships in order to push states towards democracy so that in the long-term a more peaceful world might prevail? Such a position appears untenable and Owen properly points out the pitfalls involved in actively imposing democracy in the name of international peace. He notes that the use of force or intervention to promote or protect democracy has often been undertaken by individual states in an inconsistent manner. Furthermore, as he observes, 'States generally do not do charity work: they always want power or wealth in return for intervening' (p 383). The result is that democracy becomes an excuse for self-interested power politics.

A second problem confronting democracy as a means to peace is evidenced by the violent nature of some of the new democracies, most notably Russia and South Africa. Ethnic conflicts within states have increased dramatically since the so-called victory of democracy. This leads not only to a general feeling of disillusionment with democracy particular societies but also shows the limits of

15 Owen's essay provides a substantial guide to the vast literature. See also S. Chan, 'In Search of Democratic Peace: Problems and Promise' 41 (1997) *Mershon International Studies Review* 59.

democracy in guaranteeing a more peaceful world. Marks points out that the democratic peace argument depends largely on an out-dated conception of geo-political conflicts between defined nation-states, something which just does not happen very much in the world today (p 560). Franck notes that democracy is no guarantee against the outbreak of civil war or internal conflict. But, furthermore, democracy does provide a means for a society to come together and resolve issues in a non-violent way. Decisions, negotiations and discussion can all occur without democracy '[b]ut there can never be a genuine social convergence' (p 25). The need to look at the impact that democracy has on individuals within the domestic society is crucial to the development of an international law of democracy. Ignoring broader social issues overlooks the wider role democracy has to play something that many discussions of democracy do not fully take into account.¹⁶ While this may appear self-evident, the limited procedural approach that has been taken with regard to an international law of democracy has, for the most part, not considered the impact democracy has on individuals and societies. A common assumption prevails which believes that elections bring democracy and democracy brings improvements to society, which is not always the case, underlying the need to look beyond elections in order to fully understand the development of democracy as an international legal principle.

IV

Thomas Franck was one of the first to put forward the idea of a right, or entitlement as he termed it, to democracy. His position is grounded in democratic thought and practice as defined by western industrialised nations, primarily the United States of America. He began his enquiry into democracy in international law with the US Declaration of Independence (p 1776) and from there described the development of western industrialised nations and the democracy they created as a 'hard-won tradition that most of the world now seeks to emulate' and argued that 'people almost everywhere now demand that government be validated by western-style parliamentary, multi-party democratic process.'¹⁷ In his essay here, the origins of the democratic entitlement are traced to the adoption of the Charter of the United Nations and its provisions regarding the role of the organisations in assisting former colonial territories in obtaining self-rule through the Trusteeship system (p 32). Chapters XI and XII of the UN Charter set out a process whereby a number of former colonial territories were assisted in achieving independence. Franck calls the process 'democratic' but since it often only entailed a single vote for determining the status of a territory it has also been described as 'an exercise in legitimised colonial paternalism.'¹⁸

Franck's original work developed an argument whereby the entitlement to democracy is grounded in the practice of self-determination in the international system.¹⁹ For Franck, '[s]elf-determination postulates the right of a people organised in an established territory to determine its collective political destiny in a democratic fashion and is therefore at the core of the democratic entitlement.'²⁰

16 See discussion in Part IV, below.

17 Franck, n 2 above, 48-49.

18 Gerry Simpson, 'The Diffusion of Sovereignty. Self-Determination in the Post-Colonial Age' (1996) 32 *Stanford Journal of International Law* 266.

19 Franck, n 2 above, 52.

20 *ibid.*

Franck's attempt to place the process of self-determination at the centre of the democratic entitlement must ultimately be rejected. In most of the cases he presented, the holding of a single vote determined the status of a territory, but did not guarantee the continual exercise of self-determination or an ongoing right of a people to determine their destiny. The limited number of plebiscites and referenda sponsored by the League of Nations were rarely if ever followed by regular elections or other forms of participation by individuals in their collective destiny. The same holds true in the practices of the United Nations, as many of the former colonial territories soon saw authoritarian regimes in place with a full-scale denial of self-determination, or at Roth terms it 'one man, one vote, one time' (p 508). Even under the Trusteeship system that Franck identifies as the beginning of international practice regarding democracy has, according to some critics, been accorded a greater sense of importance in the development of democracy than it actually possessed.²¹ It is difficult to sustain self-determination as exemplifying or acting as a basis for a democratic entitlement if its actual application only provides for isolated instances of territorial determination.

Franck's contribution to the present volume keeps with the notion of the democratic entitlement but offers a scaled-down treatment of democracy in the international system. The essay focuses on the holding of elections and the use of election monitoring in an effort to demonstrate how states participate in both processes in order to confirm their legitimacy. Governments will believe that they may stand as the proper representatives of a society by holding elections, and by participating in election monitoring, governments in turn recognise the validity of democracy as an international legal principle (p 29). As stated above elections alone provide only a partial picture that does not necessarily investigate any of the substantive elements involved with democracy. Franck overcomes this by viewing the election monitoring process as a means for addressing substantive issues through ensuring that elections are free and fair. Like elections, giving legitimacy to the overall process without fully investigating the consequences that may exist.²²

Fox also focuses on the existence of elections to determine whether or not international law possesses a right to political participation. For him the essence of democracy is the existence of popular sovereignty and popular sovereignty is determined by elections (p 49). He adopts a process-orientated procedural conception of participation rights with adequate participation being voting in elections even though he recognises that a wide range of rights are necessary for democracy. His work exemplifies the tendency of some to accept the sudden appearance of democracy as an international legal principle post-1990 by making use of state practice and treaty provisions that date back to 1948 or earlier. He believes that all of the major human rights conventions have established clear rights of political participation (p 90) but does not necessarily address why these rights have only been recognised recently in both rhetoric and practice. It is clear that the changing ideological environment of the international system has allowed for democracy to be discussed more openly and practised more widely. One gets the impression reading authors like Fox that international law has long embraced democracy as a legal principle even though there has only been limited international practice.

21 See Rosalyn Higgins, *Problems and Process International Law and How We Use It* (Oxford Clarendon Press, 1994) 111–112.

22 In the recent elections in Cambodia international observers were quick to declare the elections free and fair in the face of obvious irregularities, with the domestic population far from convinced that democracy in practice had actually occurred, see 'EU Loses Face for Cambodia Poll Verdict' *The Guardian* (5 August 1998) 10.

For international law, reliance upon procedure as indicative of democracy makes life easier by associating democracy with an identifiable process. If the process is carried out and deemed to be acceptable then governments which engage in the process are seen to be legitimate holders of power. The process of recognition in international law has taken on the requirement of democracy for determining the legitimacy of states and governments. Traditionally international law only required 'effective control' to be exercised over a population for a government to be seen as legitimate. How control was established or maintained was not a concern for international law.²³ The emerging requirement of democracy for recognising a legitimate government is limited to the holding of elections and the use of election monitoring. In his chapter on democratic recognition, Murphy examines a great deal of state practice in this regard only to conclude that there remains a significant degree of indeterminacy in using democracy as a criterion for recognition. Even when democracy is reduced to a procedural element other values and preferences continue to guide the process of recognition and determinations of legitimacy. This is due to the fact that recognition remains the preserve of individual states, so, even though there is talk of the importance of democracy, it is far from a guiding light that all states embrace (pp 152–153).

Tied to the issue of recognition and legitimacy is the situation whereby the democratic process of free and fair elections occurs but it produces a result that is deemed unfavourable. Recent elections in Austria have brought this issue to the forefront of international relations, as it is feared that the newly constituted government that includes the Freedom Party could lead to a throwback to the ways of the Nazi party in Germany. The fear is that, like the Nazi Party, the Freedom Party will control the government through the democratic process only to destroy or undermine the very same.²⁴ Algeria has also faced the dilemma of destroying democracy in order to save it when in 1991 the Islamic Salvation Front was poised to gain control of the government and had already pledged an end to the democratic process leading to elections being cancelled and a state of emergency declared.²⁵

Fox and Nolte dealt with this issue at length in an earlier piece²⁶ and their essay in the present volume presents the main points of their research. Democracy is about debate and solving differences through non-violent means. It depends upon the expression of opinions; as the European Court of Human Rights has stated, democracy is about pluralism, open-mindedness and tolerance.²⁷ However, Fox and Nolte believe that it ought to be permissible for a domestic political system based on democratic ideas to silence and prevent expressions that support the ending of the democratic system itself. The question they pose is how tolerant should democracies be with regard to groups who hold beliefs that are detrimental to the democratic system. They take the Rawlsian position that it is not necessary

23 See M.J. Peterson *Recognition of Governments: Legal Doctrine and State Practice, 1815–1995* (Basingstoke Macmillan, 1997)

24 On the Austrian situation see Roger Morgan, 'Austria and Europe: Rights and Wrongs' (2000) 56 *World Today* 4.

25 See Ahmed Ben Bella, 'A Time for Peace in Algeria' (1995) 51 *World Today* 268.

26 Gregory Fox and George Nolte, 'Intolerant Democracies' (1995) 35 *Harvard International Law Journal* 1. In a subsequent issue Roth and Koskenniemi responded to Fox and Nolte: see 'Responses' (1996) 37 *Harvard International Law Journal* 231. Roth and Koskenniemi's responses are reprinted in the present volume as chs 13 and 14 respectively, with Fox and Nolte's rejoinder as Ch 15.

27 *Handyside v UK*, Series A No. 24 (1979/80) para 49; *Jersild v Denmark* Series A No. 298 (1994) para 37.

to tolerate the intolerant. Rawls' belief is that if a group makes it clear that it will suppress others and not allow for any differing opinions to be expressed, then it is legitimate to suppress them (p 403). Fox and Nolte present a number of limitations to their position, the most important being an acknowledgement that international law will only allow for the suppression of democracy in order to save democracy in the most extreme cases, without clearly defining what could be considered extreme. They suggest that the better way forward would be for the underlying social problems to be addressed in order to remove support from radical parties that promise change. They also recognise that the practice of intolerance can easily lead to a great deal of abuse and indeterminacy as to what constitutes a threat to democracy (p 431).

In response to Fox and Nolte's work, both Koskeniemi and Roth raise a number of concerns, especially with regard to what is an acceptable level of suppression of perceived threats to democracy. Koskeniemi's concern is that while Fox and Nolte strive to achieve a balanced approach, there is something about their argument that shows a tendency for tolerance only to the extent that democracy is understood in established western industrialised societies. He is worried that they fail to consider the meaning of democracy held by individuals in a variety of domestic settings, who in turn are bound to have differing conceptions of democracy (p 436). Roth voices concerns along the same lines, demonstrating that the development of an international law of democracy can also easily lead to an 'Us and Them' polarisation in the conduct of international relations, with 'accepted' democracies being on the winning side and all others being labelled as lost souls that must be converted at any cost.

V

The consequence of democracy leading to an 'Us and Them' division of the international system is a major issue in the context of intervention into the affairs of other states. Interventionist action, including the use of force, to promote and protect democracy is nothing new to the international system.²⁸ Discussion on the use of force in international law to promote or protect democracy is limited by virtue of the few cases/examples that exist in modern times. The essays in this volume look to the few instances, Haiti, Sierra Leone, Grenada, and Panama, where force has been used to promote or protect democracy. Analysis of these cases is primarily limited to the doctrine established by the International Court of Justice's decision in the *Nicaragua* case.²⁹ Based on such sparse evidence, it is no surprise that differing conclusions are reached.

Reisman leads the way in justifying a democratic crusade that includes the use of force if necessary. His position on this issue has not changed since he first aired his views on the use of force in support of democracy in 1984.³⁰ He feels that international law has moved away from traditional conceptions of state sovereignty to embrace a conception of sovereignty that characterises a society where individuals are freely able to choose their leaders (p 249). He criticises a static

28 Brad Roth, 'Governmental Illegitimacy Revisited: 'Pro-Democratic' Armed Intervention in the Post-Bipolar World' (1993) 3 *Transnational Law and Contemporary Problems* 481.

29 *Case Concerning Military and Paramilitary Activities in and Against Nicaragua. (Nicaragua v United States of America)* (Judgment) ICJ Reports 1986.

30 W. Michael Reisman, 'Coercion and Self-Determination. Construing Charter Art 2(4)' (1984) 78 *American Journal of International Law* 642.

view of international law that does not look to what he feels is a great deal of progress in support of democracy (p 251). Reisman appears to believe that since international human rights have become widely accepted, democracy too has become one of constitutive elements of the international system. Any claims to the contrary are, in his view, anachronistic (p 254). Looking to progress is beneficial, but it should not be overstated in order to formulate new rules that allow for arbitrary behaviour as demonstrated by the US action in Panama in 1989.

While Reisman recognises that the use of force in support of democracy is complicated, he concludes: 'If complexity of decision is the price for increased human dignity on the planet, it is worth it' (p 258). His position on the ability, if not the duty, of states to intervene unilaterally to preserve or restore democratic governance is brought into question by Byers and Chesterman. They feel Reisman's position relies on 'simple premises and forceful rhetoric' to support unilateral action in support of democracy since multilateral efforts have not proved successful (p 262). They take the view that unilateral action is not favourable under any circumstances and that collective action is only advisable in the more extreme cases (p 291). Byers and Chesterman point out a major flaw in the thinking of those who support unilateral action as a legal response which they feel will lead to the 'Us and Them' division of international society. They express the need for supporters of the use of force to keep in mind that not every state in the world is democratic, or as democratic as some may wish. Under international law, however, regardless of the form of government every state is an accepted international actor. The belief that force may be used to ensure the existence of democracy is justified from a position of ideological superiority. This belief holds that that some international actors have a greater degree of rights than others and this in turn justifies forcible measures being used against those who do not meet accepted standards (p 270). The alternative put forth by Byers and Chesterman is to follow an accepted legal method whereby disruption of a democratic regime is deemed by the UN Security Council to be a threat to peace and security allowing for collective action to be taken to end that threat. They limit the ability to act even further by claiming that a disruption to democracy itself could rarely be seen as a threat; instead it would have to be accompanied by other events, such as refugee flows or mass killings, to constitute a threat to international peace and security (p 283).

The issue will remain contentious whether the use of force is unilateral, collective, based on specific treaty provisions (discussed by Roth) or even by invitation (discussed by Wippman). Roth points out in his discussion of the illegal nature of pro-democratic invasion pacts that 'the meaning of democracy remains inherently contestable, and never more so than in moments of actual crisis' (p 337). Placing the responsibility for democratic government on external third parties is, he feels, a complete negation of the idea of self-determination (p 342). The most worrying aspect of arguments that favour the use of force to protect or promote democracy is the division of the world into 'Us and Them'. Such attitudes are based on the idea of the victory of democracy and the belief that no other alternatives are possible – or, if they are possible, they are not tolerable.

VI

One of the more worthy attributes of this volume is that it includes contributors who are critical or more accurately, sceptical of claims of a clearly defined existence of a principle of democracy in international law. Part V contains five

essays which do not follow the optimism of Frank or the zealotry of Reisman regarding democracy as an international legal principle. What is significant about the position of the sceptics is that none of them claims that democracy is unfavourable or a bad thing for the international system or domestic society. However, they do raise a number of concerns regarding the development of democracy in international law. Much of the scepticism comes from the fact that democracy has gone from a practical non-issue in the international system (outside of the Cold War environment) to a guiding legal principle in less than ten years.

For the most part, the majority of essays outside of the section dedicated to critical perspectives give little consideration to normative issues that are central to democracy. The existence of elections and the presence of some form of external monitors to declare them free and fair are taken to be enough.³¹ Roth points out that '[i]n the exuberance of the moment, issues become conflated, and differences regarding crucial principles obscured' (p 493). Roth deals with the issue of evaluating the existence of democracy beyond the holding of elections. He views it as problematic that only a few criteria are set out as necessary so the issue of ends and goals in a particular social context are ignored, or deemed inconsequential (p 494). The majority of writers who support an international law of democracy believe that so long as an identifiable democratic procedure (elections) is being followed, the rest will work itself out in time. This is extremely problematic as struggles for democracy are pursued because individuals believe that democracy will bring certain benefits. No one desires democracy just so they can tick a ballot paper. Individuals believe that in ticking the ballot paper things will get better. Reliance upon democracy as procedure only provides a picture that is opaque at best, for it does not deal with the 'substantive social purposes for which one values the procedures' (p 496). Roth feels that for most, democracy is not just institutions and procedures but 'a normative orientation that requires creative application to the distinct problems of individual countries' (p 498). He looks to the history of democratic thought that shows democracy being much more than just procedures with attention being given to issues that are described as the 'democratic social reality' (p 502). To understand the social reality of democracy he feels that it is necessary to look at issues of substantive democracy, popular sovereignty and constitutionalism – these being the phenomena which tell us whether or not the problems facing a society have been addressed.

The same theme of the inadequacy of relying on elections is taken by Black who sees the development of democracy as providing new opportunities but also uncovers a number of new vulnerabilities that need to be addressed (p 517). Black, like Roth, sees elections as only a means, and one that does not always adequately address desired ends. Black's concern is that the emphasis on elections means various major issues facing society are left untouched. He points out three areas with specific regard to the societies of Latin America: legacies of violence, the increasing 'socio-economic disparities in society and the fact that chosen leaders have less and less influence over decisions being made in a globalising world. The crux of his position is that individuals may participate in elections that result in leaders for a society, but these leaders do not necessarily have the power or ability to pursue what society desires. He makes it clear that he is not speaking of conspiracy governance but the actual reality of who really holds power in a globalised capitalist economy. His point is that to figure out where power lies

31 Mélida Hodgson, 'When to Accept, When to Abstain. A Framework for UN Election Monitoring' (1992) *New York University Journal of International Law and Politics* 137, 146–147.

elections do not give an adequate picture, and he suggests 'following the money' to find out where real power lies (p 529).

Marks completes the collection with a sceptical review of the leading writers who have enthusiastically embraced democracy as the guiding light of the international system in such a short time frame. She takes the work of Fukuyama as a clear example of the belief democracy and market based economies are the logical progression of the development of the international system and that a victory for democracy has been assured through the collapse of communism.³² This belief she terms 'liberal millenarianism' (p 538). She goes on to examine the influence liberal millenarianism has on international law and the development of democracy as a legal principle. The belief in democracy as the end we have been striving for, and that elections are indicative of the victory of democracy, results in numerous essential issues being sidelined. Focusing on democracy as only elections overlooks the diversity of values, ideas and principles that exist in a particular society (p 558). Marks feels that belief in democracy as the end of history means that issues of citizenship, accountability and equality are not adequately dealt with (p 559). There also are issues of globalisation and the real exercise of power, and the truth about the democratic peace that are removed from the discussion (p 560). One might easily respond to such scepticism that at least elections show that movement in the right direction is taking place, so let us be satisfied without confusing the issues any further. Marks takes on this objection by pointing to the need to look more deeply at the actual impact that democracy has upon the individual and society. If the exercise of power remains essentially unchanged then the façade of democracy is no better than unmitigated repression (p 566).

VII

A major force behind the development of an international law of democracy is the activities of international organisations and the support they have given to democracy, a point made by a number of the contributors. International organisations do have a limited role in the development of an international law of democracy. As Slaughter points out they will, for the most part, 'remain servants of their member States more than masters' (p 199). In their introductory essay Fox and Roth speak of the 'limited carrots and sticks' available to international organisations when attempting to promote and protect democracy (p 8–9) but they recognise that the efforts of international organisations have been substantial. When democracy has been an issue in international law it has usually been on the basis of individual state action. This resulted in a large degree of indeterminacy and never developed into any coherent set of principles. The activities of international organisations have been and will continue to be essential to the development of an international law of democracy, providing a great deal of coherence and stability in the development of principles. However, as Schnably's essay demonstrates, international organisations will have difficulty in developing an international law of democracy, for the more active an organisation becomes in supporting democracy, the more numerous and complex the problems it faces (p 198).

International organisations, both regional and universal, will have a primary role to play in the further development of an international law of democracy.

32 Fukuyama, n 1 above

Unfortunately, international organisations have a tendency to focus on the procedural version of democracy. Their activities in a wide variety of areas such as human rights, economics and social welfare do mean they look beyond elections to determine if democracy is effective. As Crawford notes international law has pursued, and continues to pursue, a number of goals inconsistent with democracy and there remains disagreement over the details (p 115–117). This factor will limit the ability of international organisations to develop an international law of democracy beyond the idea of elections.

As Roth warns, the present approach to democracy in international law ignores behaviour beyond elections that work to further democratic values, thereby exaggerating democratic progress (lots of elections means lots of progress), resulting in an ‘unjustified triumphalism’ that is both messianic and complacent (p 494). Democracy is commonly seen as a term with no determinate content; a definitional difficulty which has not been overcome with the recent resurgence in support for democracy. Presently, democracy as an international legal principle is commonly equated to the existence of elections leading to a ‘credibility gap’ where there is a stark difference ‘between the standards [that] governments proclaim, or accept, or at least pay lip-service to, and the reality of their practice’.³³ The critical views presented in Part V make it clear that an international law of democracy concerned only with elections will not be sufficient. Elections are central to the democratic process. They are a means by which individuals and societies can express their views. By limiting democracy to a procedure, an understanding of what democracy is or should be becomes limited. It also allows for claims that democracy exists when in reality the will of the people has no impact, as individuals are not part of the processes surrounding their lives.

International law has made a great deal of progress in the improvement of human welfare, most notably through international human rights law. An international law of democracy will aid in the progress of international law only if it develops beyond elections, regardless of the difficulties inherent in taking such an approach. As an international law of democracy develops it must be premised on the ability of individuals and societies to determine their own form of social ordering and how best to pursue social needs. An international law of democracy must be constructed from the position of the individual addressing the issues as to how international law can be effective for the individual and society, not just in terms of inter-state relations. This may appear obvious but international law continues to focus on the needs and desires of states, being less concerned with non-state actors. An international law of democracy that develops according to the interests of states will remain rooted to the idea that elections automatically equate to effective democracy. An international law of democracy that develops with the needs and interests of individuals in mind will see the necessity of looking beyond elections to ensure the democracy that exist is effective and not illusory.

33 Niall MacDermot, ‘The Credibility Gap in Human Rights’ (1976) 3 *Dalhousie Law Journal* 262.

REVIEWS

Terrence Daintith and Alan Page, The Executive in the Constitution: Structure, Autonomy, and Internal Control, Oxford: Oxford University Press, 1999, xxiv 441 pp, hb £35.00.

This is a study of the central executive in the United Kingdom. The book is the outcome of research funded by the Economic and Social Research Council under the 'Whitehall' programme. The aims of the book are to explain the way in which ministers and civil servants co-ordinate and control the actions of the major departments of state. The authors intend that the analysis provided by their study will contribute to the understanding of constitutional law in the United Kingdom. The analysis is not driven by any single economic or social theory but draws generally on the writings of political scientists and constitutional lawyers. It integrates legal analysis with a detailed and technical description of the internal working of government departments. There is an emphasis on internal rules, relationships within government departments and the structure of systems of internal control. Judicial scrutiny and parliamentary control are assessed alongside internal systems of control. External audit, the work of select committees and the role of the National Audit Office are explained and evaluated.

The executive as a specialist subject has in general been neglected. This is explained by Daintith and Page (p 10) as largely the consequence of a 'preoccupation with parliamentary and judicial controls of constitutional law, rather than with the activity and organization of government that is the subject of control'. Until recently, and with notable exceptions, legal scholarship often ignored how executive powers developed. Little attention was given to the legal and constitutional basis for many of the powers currently enjoyed by the cabinet and the Prime Minister. A working presumption of legality applies to executive powers whereby it is assumed that prerogative powers authorise executive decisions provided the actions taken are not expressly forbidden by statute.

Outside the narrow confines of legal scholarship there has been some attention given to the role of the executive in the constitution. The Haldane Committee in 1918 (*Report from the Haldane Committee on the Machinery of Government* Cd.9230, 1918) identified many features of executive power, including a tendency to increase power through the growth of large departments at the cost of potential loss of effective control. In the 1930s, Sir Ivor Jennings in *Cabinet Government* offered an important case study of how executive government worked. In 1938, Keith's *The British Cabinet System* explained the importance of Executive power through the functioning of the system of cabinet committees. Richard Crossman's *Diaries of a Cabinet Minister* provided a remarkable insight into the shifting power battles in the Wilson administration. The latter dominated formal cabinet decision making through the setting up of 'kitchen cabinets' and secured prime-ministerial dominance over cabinet decisions. Such studies have their value but they fail to address the legal basis for the systems of control that the executive operates or the full constitutional dimensions of how decision making powers are made accountable to parliament.

Daintith and Page begin their study by identifying the main sources of executive power. In chapters 1 and 2 the authors set out the constitutional and legal framework. Their conclusion (p 58) is that a government department acts on the principle that the minister is responsible for the staff out of moneys provided by Parliament and may take decisions in the name of the Crown or on the Crown's behalf. Departments enjoy virtual autonomy owing to the protection afforded through 'a combination of law which vests functions in ministers; convention, which makes them responsible for those functions to Parliament; and organisation, through their membership of the central executive authority.' The courts have had a limited role in developing systems of financial control over the exercise of departmental power, mainly because of the scope of legislative authority and the protection accorded by highly permissive spending legislation (The Appropriation Act) which permits departments wide margins of error. Few cases challenge the legality of the system of financial control. The future may be markedly different, with greatly enhanced potential for litigation and judicial review under the Human Rights Act 1998.

In chapter 3 the civil service is examined as part of the internal structure involved in most of the decision making of the executive. Over the past two decades there have been cuts in the numbers of civil servants. Today, the overall number of civil servants is below 500,000. The civil service has undergone considerable internal changes. The creation of agencies and the re-structuring of government departments has taken place in an effort to increase efficiency. Formally, central control continues to operate through departments. Regulation of the civil service falls under the authority of the Minister for the Civil Service and is to be found in re-drafted codes and an increasing trend in favour of codification of internal rules and procedures. Prerogative powers are the legal foundation for most of the regulatory structure. Such codes cover terms and conditions of service, recruitment, conduct and discipline of civil servants. The impact of codification is significant: internal rules have emerged in a form that gives rise to greater transparency. Unity of purpose through common terms and conditions are thought to give some semblance of uniformity to a service that has diverse and even contradictory operations. Codification helps maintain the command and control systems that are at the centre of ministerial power as a way of directing and holding together 'a conglomerate organisation' (p 102). However significant the recent reforms may be, they fall short of the development of civil servants holding ministers directly accountable to Parliament. A permanent civil service loyal to the government of the day is seen as more beneficial than a civil service directly accountable to Parliament with the potential for the civil service to become a 'political force'. The reality is more complex, as the government of the day may effectively neutralise civil service influence through the appointment of personal advisers and the advice of the 'spin doctor'.

In locating the sources of executive power, Daintith and Page analyse the financial workings of government (chapters 4–6). There are three chapters devoted to a survey of the system of Parliamentary control of government expenditure. The technical system of control through public expenditure survey to audit is examined in some detail. A number of fundamental questions are posed and answered. How are the financial resources allocated and appropriated? What are the systems for internal control? What are the roles of the National Audit Office and the influence of the Treasury? These three chapters form a substantial part of the book. Various trends are noteworthy from reading these chapters. Codification of the various internal rules and practices has been accomplished rapidly over the past two decades. The result is recognition of the importance of technical rules and the need for increased

accountability. Transparency is more possible today than in the past. Decisions on public finance and its allocation remain department led and subject to strong systems of internal control. The overall framework of ministerial decision making remains largely unchanged, leaving large amounts of delegated authority and self-regulation in the hands of departments.

One example that exposes the systems of internal control and scrutiny is the *Pergau Dam* case ([1995] 1 All ER 611). The Permanent Secretary of the Overseas Development Administration acting as Accounting Officer had reservations about the economy and efficiency of the grant aid to Malaysia for the construction of a dam project. The Minister overruled the Accounting Officer's reservations following set Treasury procedures notified in writing in accordance with the procedures. Daintith and Page (p 172) provide a valuable analysis of the court case and how the decision to fund the dam was unlawful under The Overseas Development and Co-Operation Act 1980. The potential for the future development of judicial review in the general area of public finance is often constrained by the complex nature of financial regulation and the existence of adequate systems of internal control.

Having identified and analysed the sources of executive power, Daintith and Page turn to examine how the executive comes under the influence of legal advice. Chapters 7–11 survey the way the government organises legal work amongst government departments. The Parliamentary Counsel Office, the Crown Prosecution Service, the Serious Fraud Office and Departmental legal services are examined in some detail. In chapter 8 the power to change the law through the legislative process is examined. The drafting of legislation and the use of subordinate legislation are assessed. The influences of the European Union are considered in terms of the impact on domestic law. The European Convention on Human Rights and its impact on domestic legislation are also discussed. The manuscript for the book was completed in December 1998, allowing little opportunity for consideration of the Human Rights Act 1998. Chapter 9 provides some fascinating insights into how legal advice is co-ordinated in government and the extent to which the government's lawyers may be proactive in testing legal issues in the courts. The findings here point to a certain element of amateurism – weak structures that make co-ordination of legal advice between departments very difficult, despite recent efforts of strengthening legal co-ordination. There is also the absence of long-term strategic planning between legal advisers over concerted legal action. There is only recently an increasing awareness amongst administrators of the relevance and importance of legal advice.

The final chapter brings together the main findings of the book. The authors describe and identify many of the working practices that are found in the way the executive exercises control over departments. The trend towards codification of internal rules in the form of manuals and handbooks is an important finding of the study. Codification is also undertaken because of the influence of setting standards of conduct and recruitment. Codification strategies have provided the means of co-ordination, and through the creation of the Efficiency Unit the setting of targets and outputs. Although most of the legal authority for the regulation of the civil service is based on the prerogative with minimal statutory intervention, the executive is capable of achieving considerable control because of the adoption of a wide range of rules and regulations.

Another important finding in the book is the recognition that executive powers extend to the economic instruments of government activity, such as contract or market based relationships between the public and private sectors. Such economic

instruments have a dual purpose. On the one hand they deliver goods and services and on the other they provide a pre-eminent example of the extension of government powers. Co-ordination between departments is another example of dual use systems. Primarily intended to operate to ensure 'joined-up' government, the purpose of co-ordination is equally capable of having a centralising effect. Government policy and influence is more easily deliverable through a plethora of agencies and activities than in the past.

There are some omissions but these do not seriously detract from the book as a whole. The main focus of the book is on central government with no substantive analysis of local-central relations. There is an omission from the financial chapters on the relationship between the United Kingdom Parliament and the European Union over structural funds. The relationship between the European Court of Auditors and the United Kingdom's National Audit Office in the audit of European funds is a useful case study. Accounting for European public money amidst allegations of systemic failure, and the resignation of the European Commission, might raise the question of how the United Kingdom system of budgetary control compares with those of our European counterparts. It might be possible to evaluate the strengths of our current system and assess any European weaknesses. Attention might also be given to the potential impact of devolution in Scotland, Wales and Northern Ireland. The impact of devolution may, for example, challenge the unity of the civil service. Systems for the control of finance to the devolved governments are also useful case studies of internal systems of control. Comparisons of differences in the financial relationship between the United Kingdom's financial system of control and new devolved administrations are also useful for the study of expenditure allocations.

In assessing the value of the book from the broader perspective of the effectiveness of the controls on executive power, a number of points emerge. Ministerial responsibility remains a key element in the way internal controls are applied. Serious doubts exist as to how effective such a system of accountability may be, when so much of the control of executive power depends on self-restraint. There is also the question of the future potential for judicial scrutiny of executive decision making. Citizens will be empowered through the Human Rights Act 1998, with certain rights granted directly for the first time. This has the potential to enhance judicial power and allow judicial oversight of working manuals and codified internal rules. These may become the key documents in inspecting the standard and quality of the workmanship of the decisions of civil servants and administrators.

It may be concluded that studying the hidden workings of executive power provides a perspective from which to judge and evaluate the current Government's project for constitutional reform. The fact that so much democratic control hinges on the self-restraint of the executive is hardly new in itself. What is important is that there are myriad ways in which Parliament has stood aside, sometimes unwittingly, at other times through a mixture of inertia or impotence, to allow executive powers to fashion Parliament's own role, machinery and functions.

This leads to a thought-provoking question that arises from reading Daintith and Page's study. Is Parliament fully aware that effective parliamentary control looks less easy to imagine today when set against the reality of executive power and increased judicial scrutiny? The thesis is that executive self-restraint is essential to the underpinning of democracy and the rule of law. In a similar fashion it may be argued that a certain degree of judicial self-restraint is required in order for effective executive decision making to take place. By revealing more clearly how public power is exercised through the internal workings of executive control, this book

makes a significant contribution to the study of public law. It has an excellent bibliography, is well referenced and may be highly recommended as required reading for students of public law and civil servants interested in legal analysis of government decision making.

*John McEldowney**

A. Walters and M. Davis-White, Directors' Disqualification: Law and Practice, London, Sweet & Maxwell, 1999, lii + 500pp, hb, £60.00.

This book offers the observation that '[t]here is a strong case for saying that the law relating to disqualification of directors has become one of the most significant parts of modern company law' (para 1.01). I agree, although I would replace 'strong' with 'irrefutable'. This new text analyses the most fascinating development in modern Company Law over the past 15 years. In some senses the mere existence of the regime is an admission of failure. Well established Company Law strategies designed for the protection of creditors and the public interest (such as the capital maintenance doctrine, the ultra vires rule and the imposition by the common law of duties for directors) have failed to prevent incompetent and dishonest company directors from wreaking havoc. The only solution is to put limited liability companies beyond the reach of such individuals for a period of time.

The introduction of the revamped director disqualification regime in 1985–86 was part of a nexus of reforms designed to discourage directors of distressed companies from continuing to trade on regardless at the expense of creditors. This package offered 'sticks' (the wrongful trading sanction) and 'carrots' (the safe haven of administration). Neither of the other elements in the package has proved a resounding success. The number of cases of directors voluntarily admitting that they are incapable of rescuing a failing business and calling in professional help has been small, though a gradual rise in the numbers is apparent. The wrongful trading sanction (Insolvency Act 1986 section 214) has failed largely because it was a private law remedy restricted to a category of persons (i.e. liquidator) who by definition would be short of cash to pursue speculative civil proceedings. Herein lies the clue to the success of director disqualification mechanism to date: public expenditure in the form of the state underwriting the costs of bringing the proceedings. This is an unstable platform in view of the economy-conscious policies of modern governments and it has led to a fundamental review of the mechanism in order to make it more cost-effective. Once the Insolvency Bill 2000 takes effect, director disqualification is more likely to be pursued through the use of binding undertakings rather than formal court proceedings. Litigation in future will be reserved for cases raising genuine points of principle.

The authors approach their subject in a logical fashion. After explaining much of the background summarised above the relevant statutory provisions are unpicked. Key terminology, such as 'shadow director', is explained. Chapters 4 and 5 illuminate the central concept of 'unfitness' through discussion of formative cases and the identification of common problem areas. Much of the director disqualification litigation is concerned with procedural issues, such as time limits for action, use of evidence and costs. Walters and Davis-White devote three chapters

* School of Law, University of Warwick.

to these topics and the commentary here is excellent. The remainder of the text deals with the form and consequences of the order. The procedure under which a disqualified person may secure leave to act in respect of a named company is fully analysed in Chapter 13.

The standard of research that has gone into this work is high. All of the voluminous case law (some of which is unreported) is included together with key literature and policy documents. There is constructive usage of comparative material. The authors display an acute awareness of the implications of the Civil Procedure Rules and the Human Rights Act 1998 for their subject area. Nevertheless, the commentary will have to be read in the light of significant developments since publication, notably the analysis by the Court of Appeal of the juristic nature of disqualification proceedings in the *Westminster Property Management* case, *The Independent*, 10 February 2000 (relevant to paras 2.21 et seq) and the ruling of the European Court of Human Rights in the unreported judgment in *DC, HS and AD v UK* (39031/97) (see in particular para 5.83). The text is produced in a lucid style and contains 100 pages of appendices, the utility of which increases as they proceed. This book will undoubtedly establish itself as one of the leading director disqualification monographs in the years to come. Whether director disqualification continues to be such a fruitful area for litigation and legal writing is another matter.

David Milman*

John Collier and Vaughan Lowe, The Settlement of International Disputes: Institutions and Procedures, Oxford: Oxford University Press, 1999 xxviii + 395 pp, hb £75.00.

It is often remarked that one of the main differences between the international legal system and developed national legal systems is the absence of a centralised judicial system (possessing compulsory jurisdiction over the participants in that system) for the resolution of disputes. Since this difference is often taken to be a reflection of the weakness of the international legal system, it is curious that whilst many national legal systems are now giving greater prominence to alternative methods of dispute resolution, the international legal system is giving greater prominence to adjudication as a method of dispute settlement. It is a noticeable phenomenon that not only are there an increasing number of international tribunals, there is a significant increase in the use of existing tribunals and mechanisms such as the International Court of Justice or the ICSID arbitral tribunals. In addition, a number of recent treaties regulating important aspects of international life include interesting and innovative provisions on dispute settlement. The treaties setting up the World Trade Organisation in 1994 and the 1982 United Nations Law of the Sea Convention which came into force in 1994 are perhaps the most notable of such treaties.

In the context of these developments, there is no need to justify the publication of a book which provides an overview of the mechanisms which are available for the settlement of international disputes. Whilst there are specialist texts which deal with particular dispute settlement methods or institutions, such as the International Court of Justice or international arbitration, the publication of the book under review is particularly welcome because there are very few works which give detailed

* Centre for Law and Business, School of Law, University of Manchester.

consideration to the range of dispute settlement mechanisms used in international law. This is one of the main strengths of this book. Another noteworthy feature of this book is that, it not only considers the mechanisms available for the settlement of traditional 'State to State' disputes, but there is also detailed treatment of mechanisms available for the resolution of disputes which arise between States and non-State entities. For the purposes of this book, an international dispute is defined as a 'dispute between two States, or between a State (or its government) and a "non-State entity", such as a private commercial corporation, or an international organisation' (p. 5). There are therefore 3 useful chapters dealing with (i) international commercial arbitration; (ii) mixed arbitrations particularly the International Centre for the Settlement of Investment Disputes (ICSID) and (iii) the Iran US Claims Tribunals; the arbitral process. Whilst some consideration is given to diplomatic methods of dispute settlement, the book focuses largely on dispute settlement by adjudication – be it by standing judicial institutions or by arbitral or quasi-arbitral tribunals.

The book is divided into two parts, one dealing with 'Institutions' and the other with 'Procedure'. Part I on 'Institutions' goes through the range of international dispute settlement methods. Within that part, Chapter 2 describes the basic framework for international dispute settlement with particular emphasis being given to the diplomatic methods of dispute settlement. Chapters 3–7 then consider in detail international commercial arbitration; mixed arbitrations; the law of the sea dispute settlement system; the settlement of international economic disputes; and the International Court of Justice.

This division into parts dealing with institutions and procedure is a bit odd as the discussion of the various institutions and mechanisms of dispute settlement necessarily entails a detailed analysis of the procedures of each institution. This means that Part II on procedure only contains a single lengthy chapter that focuses on arbitral procedure. Whilst that chapter contains some material that is of relevance to other institutions, e.g. the section on admissibility of claims, most of the law described in the chapter has relevance to only arbitral proceedings. Another unfortunate consequence of this division of the book into parts dealing with institutions and with procedure, is that the division does not facilitate detailed treatment of particular issues in the context of particular institutions. To take one example, whilst Chapter 8 on the arbitral process contains a section dealing generally with the applicable law in international arbitrations, one would have expected that Chapter 4 (dealing with ICSID arbitration) would deal with the particular problems regarding the applicable law in ICSID arbitrations which arise from Article 42 of the ICSID Convention. Unfortunately this is not the case.

The best chapters in the book are Chapters 7 and 8 dealing, respectively, with the International Court of Justice and the Arbitral Process. Both of these chapters are comprehensive and provide much more than an introduction to the ICJ and to international arbitration. Not only do these chapters provide an excellent starting point for anyone wishing to understand how international adjudication works, they highlight many of the difficult issues which arise in proceedings before these tribunals and how these issues are solved. For example, Chapter 7 includes detailed consideration of the jurisdiction of the ICJ and contains an excellent section analysing how the Court has dealt with the reservations most commonly made to the ICJ Optional Clause declarations. Likewise Chapter 8 deals with the various issues which arise in arbitral proceedings from submission of the dispute to arbitration right through to challenges to arbitral awards. In this chapter, as well as in Chapter 3 (International Commercial Arbitration), the relationship between and the importance

of different sources of arbitral law and procedure is emphasised by references to treaties, national laws and arbitral rules developed by various institutions.

One weakness of this book is the imbalance between the chapters. Whilst Chapters 7 and 8 are very substantial indeed – containing 62 and 85 pages respectively – the chapters dealing with other dispute settlement mechanisms are not as comprehensive. Chapter 5 dealing with dispute settlement system provided for in the Law of the Sea Convention contains a good description of that system but is rather brief and does not explore many of the interesting issues which are likely to arise under that system. For example, there is no exploration of whether the International Tribunal for the Law of the Sea may deal with non-law of the sea disputes. Similarly, whilst Chapter 6 gives a good overview of the different systems of dispute settlement provided for in treaties regulating international economic affairs, its treatment of the WTO dispute settlement system is disappointingly brief. The book does not do justice to the importance of the WTO dispute settlement system, the range of procedural issues that arise under that system and the rich case-law that has which have raised and attempted to solve these problems.

The preface to this book indicates that it was conceived by the authors as an introductory text to be used in teaching. The authors have accomplished more than they set out to do. There is no doubt that the book will be particularly useful for students taking courses in international dispute settlement. In this regard, it is pleasing to note that OUP are publishing the book in paperback and at a price affordable to students. However, scholars and practitioners seeking to get an overview of the diverse dispute settlement methods currently used in international law will find this book to be a most useful addition to their libraries.

Dapo Akande*

Jürgen G. Backhaus (ed.), The Elgar Companion to Law and Economics, Cheltenham: Edward Elgar, 1999, xii + 548pp, hb £125.00.

Boudewijn Bouckaert and Gerrit de Geest (eds.), Encyclopedia of Law and Economics. Volume I: The History and Methodology of Law and Economics. Volume II: Civil Law and Economics. Volume III: The Regulation of Contracts. Volume IV: The Economics of Public and Tax Law. Volume V: The Economics of Crime and Regulation, Cheltenham: Edward Elgar, 2000, xvi + 1094 pp (vol. 1), xv + 807 pp (vol. 2), xvi + 1205 pp (vol. 3), xiv + 365 pp (vol. 4), xv + 722 pp (vol. 5), hb £865.00.

These two collections obviously are aimed at the library market rather than individual purchasers. *The Elgar Companion to Law and Economics* is the best value and possibly the most useful, containing as it does some very good overview essays on primary law and economics themes. There are some strange omissions: the volume contains next to nothing on either tort law or the economics of crime; furthermore, the lengthy section on classical authors in law and economics overlooks certain of those, such as Aaron Director and Herbert Simon, who played a crucial role in the development of law and economics as a sub-discipline, yet highlights the contributions of Rudolf von Jhering and Thorstein Veblen – figures who can at

* Department of Law, University of Durham.

best only be described as on the periphery of law and economics. The quality of the contributions is, none the less, consistently high and anyone researching in law and economics will find this volume to be a very useful resource.

The *Encyclopedia of Law and Economics* contains superb bibliographies and essays on different areas and schools of law and economics research, on applications and critiques of economic analysis of law and on the development of law and economics throughout various jurisdictions. Refreshingly, the collection emphasises neither Chicago nor even the United States. One imagines that many libraries, alas, will not be able to afford it.

Neil Duxbury*

K. D. Ewing and C. A. Gearty, *The Struggle for Civil Liberties: Political Freedom and the Rule of Law in Britain, 1914–1945*, Oxford: Clarendon Press, 2000, xvii + 451, hb £50.00

A decade after having catalogued the assault upon civil liberties in the 1980s in *Freedom under Thatcher: Civil Liberties in Modern Britain* (Oxford: Clarendon Press 1990) Ewing and Gearty have turned their attention to the first half of the 20th century. This was a period when governments and their agents feared for the safety of the state because of war and the real and imagined threat of subversion. There are six topics whose civil liberties' aspects are studied: the First World War; the Communist Party of Great Britain case; the General Strike and its aftermath; the National Unemployed Workers Movement, the rise and fall of the Fascists and the position in Ireland before partition and subsequently in Northern Ireland.

The criterion which is used for assessment is the Rule of Law. Cue the standard demolition of Dicey's formulation and its replacement with a more politically neutral conception of legality with the properties that government should have legal authority for its actions; such legal authority should be clear, published in advance and prospective; that it should not provide indemnities for government officials, nor be used to confer unduly wide discretionary powers. This conception also requires an autonomous accountable legal process in which autonomy refers to judicial independence in the sense of a separation of powers, with a separately constituted judicial branch exercising the judicial power.

Civil liberties are distinguished from human rights. Drawing on an essay by Marx, civil liberties are connected with political participation and include the right to vote as well as freedoms of association, assembly and expression. Human rights protect the autonomy of the individual, for example a person's privacy, conscience and property.

The conclusions they reach are that the different branches of government were not fully committed to the legality principle. While legal authority for action might be sought by the executive through legislation, this often took the form of broad enabling statutes such as The Defence of the Realm Acts 1914–15, the Emergency Powers Act 1920, the Restoration of Order in Ireland 1920 and the Civil Authorities (Special Powers) Acts (Northern Ireland) 1922, 1928, 1933 which permitted the making of regulations some which were draconian. The courts as the makers of the common law, with its claim that it protects the citizen, did not live up to this rhetoric

* School of Law, University of Manchester

in their decisions. The executive might, understandably, be given some leeway in war time but the courts were too accommodating. Take for example, *R v Halliday, ex parte Zadig* [1917] AC 260, in which only Lord Shaw of Dunfermline protested that it was wrong to interpret the Defence of the Realm Act as providing the lawful basis for internment without trial. Incidentally this was a much more robust statement than the more famous dissent of Lord Atkin in *Liversidge v Anderson* [1942] AC 206.

After the war some of the emergency legislation was incorporated into ordinary law and then the courts authorised practices previously thought dubious. For example, the police were allowed to seize property by reference to subsequent justification in *Elias v Pasmore* [1934] 2 KB 164, and to restrict public meetings by giving discretion to police officers which makes reference to previous and not just current circumstances in *Duncan v Jones* [1936] 1 KB 218. Decisions such as these effectively gave the police the power to control public meetings and this tended to be exercised in a way which narrowed the range of political views being expressed. Compare the hostility towards the communists and National Unemployed Workers Movement with the initial toleration of the British Union of Fascists.

Parliament does a little better in their judgement. Whilst authority was given for the making of regulations some of which were not required to be approved by Parliament, there was some element of accountability. The House of Lords secured amendments to Defence of the Realm regulations which would have subjected British citizens to trial in courts-martial rather than the civil courts for offences where the penalty could be capital punishment. During the General Strike the Opposition did seem to win the argument that there was no need for an eighth Proclamation of Emergency required to justify regulations under the Emergency Powers Act 1920. The Home Secretary may have been bested in debate but the parliamentary oversight did not stop the government prevailing. The Incitement to Disaffection Bill introduced in 1934 was improved and the Public Order Act 1936 is adjudged to have been a measure which took appropriate action on a particular problem. Parliament is also criticised for acquiescing to the creation of a constitutional convention of not interfering in matters devolved to the Northern Ireland such as the taking of wide-ranging discretionary powers under the Civil Authorities (Special Powers) Acts (Northern Ireland) and the substitution of proportional representation with first past the post.

While Ewing and Gearty suggest that this historical study will have contemporary resonance, they do not spell out their views. They certainly do not put their faith in the judiciary. Is it a vain hope that those who are now appointed to the bench are less executive-minded than many of their predecessors and that this, added to their human rights training and responsibilities will mean that they do live up to the common law rhetoric? Two recent cases might suggest greater protection for civil liberties. First the House of Lords in *DPP v Jones* [1999] 2 AC 260 permitted peaceful and non-obstructive meetings on the highway, albeit by a three to two majority and the dicta of Judge LJ in *The Guardian and The Observer* cases in the Court of Appeal on 21 July 2000. In this case Judge LJ did not rely upon the European Convention on Human Rights but said that the protection of press freedom as part of freedom of expression and the right to non-discrimination were 'bred in the bone of the common law'. Time will tell if they will be typical or not of the judicial approach to the protection of civil liberties. One change for the better is that former politicians, and the law officers in particular, are no longer being appointed to the bench.

If Parliament is to play its part then the modernisation process still has some way to go before it can improve its holding of the executive to account. The proposals in the

Liaison's April 2000 committee report to shift the balance towards Parliament were not accepted and if the Conservatives were to form a government it is unlikely that they would implement all of their internal commission's report *Strengthening Parliament*. Governments do not easily give up power, and they usually take very seriously the security of the state so that they continue to push legislation to protect it, such as the Terrorism Act 2000 and the Regulation of Investigatory Powers Act 2000. The former statute continues the practice of putting former previous emergency legislation into ordinary law and the latter statute signifies that the promotion of the United Kingdom as a leader in the development of e-commerce cannot overcome the ingrained reflexes of the security state. Only a few countries have imposed arrangements to seek to get at encrypted material and it remains to be seen if internet service providers relocate their servers to those states without such controls in order to retain their clients. These statutes will surely provide tests which will indicate the courts' support of civil liberties and no doubt in another decade we can read Ewing and Gearty's assessment of it.

Brian Thompson*

Janet Dine, The Governance of Corporate Groups, Cambridge: Cambridge University Press, 2000, xvii + 208 pp, hb, £37.50.

Globalisation by transnationals in a context of free market economics has potentially tragic consequences exemplified by polarisation of income and environmental damage. If the activities of giant transnational corporate enterprises are not controlled the result may be systemic collapse. The central aims of this important book are, first, to present an explanation of this situation and, secondly, to suggest a legal framework that may assist in reversing the trends observed.

The book begins with a discussion of the theoretical underpinnings of companies and their governance. The first chapter makes the points that the models of companies that have been adopted in various jurisdictions are shaped by the theories concerning the place of companies within society and that such theories shape the relationship that companies have with all the participants in their economic activity and with their regulators. Dine focuses on the three most influential theories – the contractual, the communitaire and the concessionary theories – to show how company law has been developed in the UK and the USA. Adherence to free market contractualist theory has resulted in the lack of a sufficiently strong level of regulatory control of corporate activity. Dine concludes that the UK is beginning to see the company not as a contract made between owners for their own profit but as a commercial enterprise with purposes of its own, which must be furthered by decisions taken with its interests in mind rather than the immediate and narrow property interests of its shareholders. However, this move from contract to constituency models has created a crisis in corporate governance. The majority decision of the shareholders no longer acts as a convenient yardstick, and yet both legislation and codes seem to depend only on shareholders as providing protection against poor or fraudulent management. Not only does this crisis affect single companies, but its impact is just as great for corporate groups. Indeed, theoretical underpinnings will also influence the structure and behaviour of groups.

* Liverpool Law School.

Dine moves on to discuss comparative perspectives of the governance of groups. Noting first the lack of debate on the regulation of groups, she presents some definitions of groups and explores the role of multinational and transnational groups, highlighting the increasing levels of activity on an international basis. The implications of this increase of activity by such companies are potentially enormous. Multinationals and transnationals exercise significant power and where they are rooted in free market theories they rely upon arguments in support of profit maximisation to justify the exercise of such power. This can lead to fraudulent or immoral use of the corporate veil to shift resources between companies to defeat outside interests; the capacity for oppression of minority interests; and to difficulties in defining a single economic unit in a changing environment. All such approaches can enable groups of companies to avoid state regulation. Private international law has not responded in a coherent fashion to the problems raised by transnational companies, in part because there exist conflicting legitimate expectations that are supposed to be protected by conflicts rules. One solution, suggests Dine, is to adopt an amended version of the *siège réel* doctrine, not to determine recognition of a company but to judge the validity of its decision-making according to its constitution. In this way, validity of decision-making would be reformed not to suit the law of the place of incorporation but to the suit law where central decision-making actually took place or the law of the place where the decision had its most significant impact.

A regulatory framework is required, but how is this to be shaped? It is clear that theoretical underpinnings both provide justifications for and influence the methods or models of regulation. Returning to the theories identified at the beginning of the book, Dine demonstrates how they shape the regulatory framework. For example, contractual theories call for minimum interference whereas concession and communitaire theories make companies open to state regulation and permit the use of companies as direct social engineering tools or sources of distributive justice. Dine observes that the free market contractualist approach has been a significant factor in formulating external prescriptive rules, with the result of creating ineffective marginalised state regulation that excludes regulators from the heart of corporate decision-making. She suggests that there is necessity for internal controls. Guideline codes, internal regulations and an outside authority must be found, as well as ensuring that both the state and affected individuals can enforce adherence to the dual concessions that have permitted the company to operate.

Dine continues her argument by demonstrating the urgency of creating effective legal mechanisms for challenging corporate decisions. She provides vivid examples to support her thesis that transnational companies are out of control. Not only do the largest companies have enormous financial power, but they are able to wield such power to divert attention from their production of environmental damage, to displace indigenous peoples by exploiting natural resources, to dominate small producers, and to increase the globalisation of poverty. Existing regulatory mechanisms have been starkly ineffective in curbing these devastating trends. As Dine argues, 'the political and economic power of many transnationals dwarfs the traditional regulator, the nation state, makes the idea of shareholder control laughable, and leaves us groping for alternative control mechanisms.'

Dine's suggested alternative includes a participatory model that adopts the dual concession theory; accepting that the legitimisation of corporate activity rests with the communities where corporations operate, not merely with the state that has 'delegated' powers to them. Legitimacy of control lies with the state delegation of the power to operate with limited liability and the devolution of powers by the founders, on condition that the norms of the local community are adhered to. Such

communities must be equipped to challenge the global companies. To create a legal framework to assist these communities it is necessary to think both locally and globally, requiring the consultation of public interest groups in the community and that nation states be able to refer to international law to create internal rules to enforce adherence by the global companies to the local community's requirements. Dine introduces the concept of 'constitutional disability' which allows challenge to a corporation's decisions where there has been a lack of proper consultation. Constitutional disability in the organs of a company will mean that acts taken in ignorance or in contravention of the disability are an abuse of the powers of the company and, as such, are a breach of fiduciary duties not susceptible to ratification by shareholders. Alternatively, one could focus on the need for proper systems of decision-making as a guide to the enforcement of fiduciary duties, so that if proper systems of management are not in place the company will be damaged.

This book is an important contribution to the literature on corporate governance. It is of relevance to company lawyers and environmental lawyers and would certainly interest anyone concerned with global issues. It turns attention towards the increasingly powerful transnational corporate groups and makes clear the significance of theoretical underpinnings for shaping the regulatory structures required to control the activities of companies. The suggested way forward might be criticised by some readers for appearing to be too much of a compromise since it emphasises procedures – constitutional disability or proper systems of decision-making – possibly at the expense of substantive changes to corporate structures or radically different theoretical underpinnings. However, Dine at least opens up the debate. The suggestion she makes should provoke new thoughts.

The book is eloquent and passionate and therefore extremely readable. At times its message is disturbing, but Dine intends it to be so. It urges the reader to consider ways of challenging corporate decisions more effectively. Dine states in the final paragraph of her book that 'it is incumbent on those of us who have the resources to think and write about the questions' with which she is concerned. The book succeeds in making the contribution that Dine hopes for and it is likely to inspire further contributions to the debate from those who read it.

*Charlotte Villiers**

Hugh Collins, Regulating Contracts, Oxford: Oxford University Press, 1999 xv + 386pp, hb £50.00.

This is a very ambitious piece of scholarship, employing an interdisciplinary approach to the analysis of contract and contractual relationships. Describing his methodology as a 'tripartite conversation between sociology, economics and law', Collins's approach requires a detailed and thorough exposition of the nature, operation and practices of contractual relationships, in addition to reflecting upon the effect that law has on such relationships.

Collins is primarily concerned not with enforceability of contracts but with whether contracts are the best way to organise arrangements between parties. Part of Collins's argument is that, to a large extent, the current contractual regime is irrelevant in so far as technical breaches of contractual duties do not necessarily lead

* School of Law, University of Glasgow.

to corresponding contractual rights being enforced. Furthermore, contractual disputes are increasingly being handled by resort to alternative methods of dispute resolution and, in many industries, self-regulatory practices. His inventive analysis categorises all types of intervention in markets as being types of regulation. By so doing he is able to investigate the 'comparative advantages of different regulatory strategies with respect to the objectives of the legal regulation of contracts'. Collins's point in the conclusion, that 'disintegration' within contract will be the upshot of a quest for uniform codes across national boundaries with regard to particular types of transaction, is particularly illuminating. Similarly, his arguments concerning the need for new techniques within contractual regulation, and particularly his claim that certain types of contract will necessitate specific forms of regulation (ie, areas such as standard form contracts), provide much food for thought.

Collins's approach is a bold and imaginative one. That he succeeds in merging the often conflicting discourses of economics, sociology and law is a tribute to his analytical expertise and grasp of often difficult concepts. In addition, his historical understanding of the growth and development of contract law allows him both to excavate and to re-appraise key areas of contract law to support his arguments. This book will provide the basis for much future research and encourage a wider approach to the study of contracts. It is a highly original and important work that will be of immense use to academics working within the field of private law. In addition, whilst this work is far from a traditional student text, those studying contract will find that his analyses add a fresh and invigorating perspective to their appreciation of the subject.

Guy Osborn*

Lucien Karpik, French Lawyers A Study in Collective Action, 1274–1994, Oxford: Clarendon Press, 1999, xii + 329pp, hb £55.00.

Lawyers in England will find this to be interesting reading. Initially published in French with the title *Les Avocats: l'État, le public et le marché, XIIIe-XXe siècle* (Editions Gallimard, Paris, 1995), the book is the work of Lucien Karpik, a distinguished French sociologist from the *École des Mines* in Paris. Karpik looks at the history of the legal profession in France and, in particular, at the profession's specific interests, quest for autonomy and mobilisation around public issues. The book covers over seven hundred years with both discipline and grace. It puts to a test a number of commonly-held views about the legal profession as a whole – for example, the myth of continuity or the so-called '*exception française*' (whether the French legal profession was ever free from political influence). The book's intellectual matrix is Weberian – it looks at how the legal profession as a collective body will "do things" depending on the specific meanings lawyers may or may not attribute to what they do as individual members of society. In one popular application of this approach, then, an organized legal profession must be analysed in terms of whatever attempts lawyers might have made to prevent exploitation from inside (for example, corruption) but also to prevent control of the legal profession from outside. 'In this perspective, and since everything revolves around conformity and deviation, the profession exists only through the mechanisms which ensure the reproduction of

* School of Law, University of Westminster.

its culture and social structure, and thus maintain its internal equilibrium'. An alternative 'Weberian' approach, on the other hand, has led scholars to view the legal profession in terms of the monopoly lawyers have over legal services. On this account, the privileges enjoyed by lawyers would be simply 'the counterpart of their capacity to exclude competitors'.

Karpik seeks to mediate between those two popular views of the legal profession by paying attention to the role of politics. As the French title of the book suggests, he is able to distinguish no fewer than 'three' legal professions which have succeeded one another in the course of the centuries: the State bar, the classical bar, and the liberal bar. These 'three' legal professions have three different organisational references – the State, the public and the market. Especially instructive are the chapters on the classical bar, which show the French legal profession's march into modernity upon its gradual reorganisation (from the second half of the 17th century onwards) around the concept of 'independence'. By 'independence' Karpik means 'the capacity of self-government' and 'the capacity to define and implement a common strategy'. Another concept around which the classical bar would reorganise itself is that of 'political liberalism' by which the author means the support given by the legal profession to an emerging moderate State as well as to a more expansive civil rights agenda.

If, however, the book's overall project is convincing, its strong points also betray some weaknesses. The book says virtually nothing about the 'professionalisation' of the legal profession – how, specifically, French lawyers came to be seen to exercise a socially higher-ranking '*profession*' rather than a '*métier*'. Another problem is that the second part of the book (entitled 'Today') would probably appeal sociologists interested in law much more than to lawyers interested in sociology. By contrast, the last few chapters of the book are unlikely to fail to interest those who are keen to grasp the specific complexities of the French (or, perhaps, Parisian) legal profession of today. On balance, then, this is an excellent, highly readable book that will go a long way to dispel a number of misconceptions that both common and civil lawyers continue to display about the French legal tradition.

Umberto-Igor A. Stramignoni*

Anthony Aust, Modern Treaty Law and Practice, Cambridge: Cambridge University Press, 2000, xxxix +443pp, hb £75.00, pb £27.50.

Treaties are the factotums of international law, but despite the many uses to which treaties can be put, customary international law has been more frequently the subject of academic inquiry. Perhaps the attraction of customary international law for academics is that it provides writers with the opportunity to have an effect upon the corpus of international law. For authors in States which automatically incorporate customary international law into their domestic legal system, the customary status of a rule of international law may transform it into one of national law. Thus, by arguing in favour of the customary status of a rule of international law, writers can attempt to co-opt the municipal enforcement structures of a State.

To assert that customary international law has taken more than its share of academic attention is not to say that treaties have been ignored; indeed, some of the

* London School of Economics.

most well respected literature in international law has been dedicated to treaties. Lord McNair's *The Law of Treaties* immediately springs to mind as an example. With books like McNair's and Sir Ian Sinclair's *The Vienna Convention on the Law of Treaties* as standard works in the area, any new work on treaties has to live up to the considerable reputations of its predecessors or suffer in comparison with them. Aust's book continues the tradition of high quality literature on this source of international law, and deserves a place in the library of any academic or practitioner interested in treaties.

The book itself is based primarily around the 1969 Vienna Convention on the Law of Treaties, the modern code of international law governing the creation, interpretation and termination of treaties. Owing to the status of the Vienna Convention as customary law in its own right this focus is the right one and provides a sensible structure for the volume.

The book begins with an evaluation of the Vienna Convention (chapter 1), before moving on to examine the concept of a treaty, and its informal cousin, the memorandum of understanding (chapters 2–3). All the major issues relating to the law of treaties dealt with by the Vienna Convention, such as reservations (chapter 8), interpretation (chapter 13) and the effects of treaties on third parties (chapter 14) are discussed in a helpful fashion. Aust also includes an enlightening discussion of the complex issue of succession to treaties, which is dealt with by the 1978 Vienna Convention on the Succession of States in Respect to Treaties. As examples of practice in this area he uses the divided and reunified Germany, the former USSR, the former Yugoslav republics, what was Czechoslovakia and the European Community (pp 311–320). All the examples are apposite, and his examination of the various cases emphasises the pragmatic aspects of succession.

As a whole Aust takes a very practical approach to the subject, reflecting his immense experience in the day-to-day negotiation, interpretation and application of treaties in the Foreign and Commonwealth Office. Particularly noteworthy in this regard is chapter 16, where Aust draws on his practical experience to give an incisive account of the responsibilities of a State acting as a depository of a treaty, and the pitfalls which accompany that role. Special mention must also be made of chapter 23, which deals with the drafting of treaties. A number of useful practical recommendations are made in relation to the actual processes of negotiating and drafting. For example, Aust makes eminently sensible suggestions on the margins and spacing to be used on drafts of treaty articles (p 360).

Aust's disavowal of any scholarly pretensions for the book (p 2) is in some ways misleading. This book contains sophisticated discussion of certain aspects of treaty law (such as the distinction between memoranda of understanding and treaties (pp 41–44)), although other sections, such as the one on the immensely interesting and controversial concept of *jus cogens* are rather short (pp 257–258). This, again, reflects the practical nature of the work.

It appears that Aust's intention in writing this book was to provide a practitioners' manual of sorts on the law of treaties. His writing is bright and lucid, on occasion bringing humour to a subject that is often seen as rather dry. This is an extremely valuable book, which deserves to be read by academics and practitioners, and particularly by those about to embark on treaty based dealings, or who wish to see how the law of treaties is applied in practice.

Robert Cryer*

* School of Law, University of Manchester.

Costas Douzinas, *The End of Human Rights: Critical Legal Thought at the Turn of the Century* Oxford: Hart Publishing, 2000, ix + 410pp, pb £15.00.

This book examines one of the most powerful forces in contemporary political practice: the discourse of human rights. Following a conventional format, the book is divided into two parts: in the first part, Douzinas traces the history of human rights from classical Greek origins to its modern practice, whilst the second part contains studies of the philosophy of human rights as revealed by such thinkers as Burke, Marx, Kant, Heidegger and Freud. The book contains perceptive analyses of aspects of the theory and practice of human rights, thereby ensuring that it will become required reading for scholars of the subject. But as a project it does not quite succeed. And the main reason for this is that it remains unclear whether it is intended to be a monograph or a textbook.

Douzinas suggests that *The End of Human Rights* can be viewed as 'a textbook for the critical mind and the fiery heart' (p 4). As a textbook it is hardly orthodox, though it does include useful studies extending from the intrinsic link between natural right and justice in classical natural law to the contribution which psychoanalysis can make to a modern understanding of right. However, as its title suggests, the book also promotes an interesting thesis. Douzinas argues that in late-modern times the language of human rights has been transformed from a discourse of rebellion and dissent into a criterion of state legitimacy and a new type of positive law. But once human rights become hijacked by governments and bureaucrats, once they become instrumentalised and lose their transcendent character, then we have reached a period marked not by the triumph of human rights but by their demise. The problem is that this thesis is not rigorously pursued. A number of the chapters, especially in the second part of the book, do not directly touch on this theme and consequently detract from its development. This is unfortunate, especially since it permits Douzinas, who initially roots his thesis in a radicalised version of natural law, to slide into a rather opaque position, one which simply celebrates the paradoxical and aporetic.

The strength of Douzinas's natural law thesis is that, by using the natural as a criterion of the good, he is able to mount an effective critique of the narcissism of contemporary society, "in which every desire is a potential right it is forbidden to forbid" (p 319). But that thesis has its difficulties, since a politics geared to the natural order appears to be intrinsically inegalitarian. Although acknowledging the influence of Leo Strauss, Douzinas denies that his thesis returns us to an anti-democratic form of classical thought. However, his attempt to present the case for a modern version of humanism that is neither historicist nor metaphysical avoids the major problems and then shrouds itself in the jargon of post-modernism. Had he worked through his thesis systematically, Douzinas might have produced something leaner, more-focused and much more powerful than this stimulating, insightful and rather unorthodox textbook.

*Martin Loughlin**

* Law Department, London School of Economics & Political Science.

Roger Cotterrell, *Emile Durkheim: Law in a Moral Domain*, Edinburgh: Edinburgh University Press, 276pp, pb £16.95.

When I was a young academic in the 1970s I taught Sociology of Law around the classic trinity of sociological giants, Durkheim, Weber and Marx. It was a reading course and the choice of texts for Weber and Marx posed a challenge. But Durkheim was easy. His ideas on law were laid out in a single text, *The Division of Labour*. So the class read this and discussed a simple theory of legal evolution from traditional societies to modern societies. Traditional societies, according to our reading, had undifferentiated social roles and modern societies had a high degree of specialisation. Each type of society had a different type of law. Traditional societies had repressive law and modern societies had restitutive law. This simple and easily criticised theorist was easily the least impressive of the three authors. A benign neglect of Durkheim's ideas in my teaching followed, shared by most other sociologists of law. While still getting his due as one of the founding fathers, he has been seen as a much less sophisticated theorist than Weber and Marx.

It is the great virtue of Roger Cotterrell's new monograph on Durkheim that he challenges this view fundamentally. He argues that *The Division of Labour* presents a flawed view of the sociological role of law, one that Durkheim himself recognised as inadequate and implicitly rejected in his later work. The key to a proper understanding of Durkheim, Cotterrell proposes, is his view that legal rules are moral rules.

In *The Division of Labour*, Durkheim argued that in pre-modern societies law and morality were expressions of a common collective consciousness. There was a shared focus of belief based on religion. Law was a reflection of these beliefs. It was the alternative basis for the moral foundation of society in modern secular societies that caused Durkheim intellectual problems. In *The Division of Labour* (1893) he argued that the communal foundation of modern society was the mutual interdependence created by the division of labour. In his post-1895 writings, Cotterrell explains, this mutual interdependence is forgotten and in its place is substituted 'collective representations' of shared perceptions and common understandings. These are the moral foundation of society and are expressed through law. Law reflects general beliefs of what rules there should be as to what is right and wrong.

This epistemological shift by Durkheim has important consequences. After 1895 he simply ceased to use the terms 'repressive law' and 'restitutive law'. This absence undermines his thesis of legal evolution, which describes a transition from repressive to restitutive law. Cotterrell calls this thesis the most famous aspect of Durkheim's legal theory but also the most frequently misunderstood. On Cotterrell's reading, this is not an empirical hypothesis about law but a sociological inquiry into the moral essence of complex societies. The historical interest is in the changing moral ideas that constitute the content of law. For Durkheim, the empirical refutation of his thesis would be that relations of reciprocity and not communal religious beliefs were the moral foundation of pre-modern societies. The discarding of the concepts of repressive and restitutive law also results in the rejection of the related thesis that law is an index of social solidarity. Durkheim no longer appears as a crude empiricist who counted the number of legal rules in a society to measure the intensity of social interaction.

Cotterrell paints a different and much more interesting picture of Durkheim's contribution to legal theory. This centres on law having a moral meaning and seeing

legal rules as a subspecies of moral rules. The formalist separation of law and morality, in which modern law is seen as a rational non-moral phenomenon, is unthinkable for Durkheim, for whom law and morality are thoroughly intertwined. But this does not mean that Durkheim is some kind of natural lawyer. The shared beliefs of society as the moral foundations of society are facts that can be discovered by empirical methods. They are relative, not absolute, truths that depend on their historical and social context.

What, then, for Durkheim is the moral foundation of modern law if it is not mutual interdependence? Cotterrell argues that it is individualism; not as egotistical self-seeking but as the communal value of mutual respect for the freedom and dignity of others. This is the value that grounds modern concepts of legal rights and social responsibilities. For Durkheim, this moral imperative gives the individual a personal commitment to law and makes law meaningful, authoritative and worthy of respect. It is the sentiment that substitutes for religion in modern societies.

This is an important reassessment of Durkheim's views on law. It is based on a wide and careful reading of all of Durkheim's texts. But it is not an uncritical reading. Cotterrell's central criticism is that Durkheim is a legal positivist for whom a legal rule is what it is, an objective observable fact. This leaves Durkheim unable to grasp the concept of law as an interpretative act with a subjective meaning for those actors who obey it. Another limitation is that Durkheim cannot see law as either a process of dispute resolution between private interests or as the instrument of particular groups within society. Either perspective would undermine his view that law is the instrument of society as a whole and the expression of common beliefs. Durkheim consequently neglects the function of law as a compromise between opposing economic interests. Cotterrell further adds that viewing law as an expression of communal moral values leads to an inadequate concept of the State. For Durkheim, the State is a collective voice that authorises popular morality by translating underlying values into specific laws. It is the brain of society, which learns about society's moral identity and then expresses it as individual legal rights. Cotterrell argues that this ignores an autonomous 'governmental morality' which gives the State a management role to co-ordinate effective social interdependence.

The prime virtue of Cotterrell's monograph is that replaces the picture of Durkheim as a legal evolutionist and crude empiricist with one of a more sophisticated and consistent inquirer into the moral foundations of law. No longer can explanations of Durkheim's sociology of law concentrate solely on *The Division of Labour*; indeed, Cotterrell makes a convincing case for treating it as a false start to be ignored.

Cotterrell, however, is a theoretical relativist. All legal theories offer only partial truths. Theories are just different ways of seeing but they never reveal the whole picture. This makes him too sympathetic to Durkheim's relativism in which moral values are empirical and non-judgmental facts. Durkheim has no answer, for example, to one of legal theory's oldest conundrums – is there a duty to obey immoral laws even if they express a collective sentiment? If law is only a fact, then we do not need law as interpretation. If we all agree that an act is wrong or immoral, we do not need a legal rule to tell us. But if we disagree as to the morality of an act, we need law to adjudicate as to whose morality is right. This is a clash of fundamental values, not an interpretive difference as to the external fact. When law adjudicates between values, it cannot be expressing a single moral consensus. To understand morality in law we need to understand how and why legal actors choose between different moralities. This needs a theory of power and an

explanation of how law balances values. Despite Cotterrell's rehabilitation of Durkheim, he remains an inadequate theorist of power compared with either Weber or Marx.

*Ken Foster**

Richard L. Revesz, Philippe Sands and Richard B. Stewart (eds), Environmental Law, the Economy, and Sustainable Development, Cambridge: Cambridge University Press, 2000, ix, + 436pp, hb £40.00.

In recent years environmental policy has been challenged by two somewhat incongruous sets of ideas. Since the late 1980s, 'sustainable development' has been promoted by scholars and policy makers alike as the new normative basis of environmental law. The principles of sustainability emerged from European and broader international networks and are based on defining economic prosperity as dependent on environmental health. The rise of sustainability thinking has been paralleled by the emergence of a rival discourse, often known as 'economic rationalism', which has increasingly become staple government policy in many Western jurisdictions. Economic rationalism asserts the superiority of market mechanisms over administrative regulation, and calls for the reassertion of economic principles as the basis for public policy. Much environmental regulation has become a target of economic rationalist criticism, and consequently significant tensions have erupted over how to promote long-term environmental protection in the midst of economic deregulation reforms which tend to diminish the role of the state in resource management decision-making.

Themes of juridification, regulatory overload, financial crisis and democratic deficit have been rallying points for critiques of command-style environmental law, along with other areas of public regulation. Increasingly, these critiques have expanded beyond the traditional focus on national jurisdictions, to practices and problems in a regional and transnational setting. A common view is that the methodologies of environmental regulation should be better aligned with the imperatives of a deregulated economy so that environmental policy is achieved more efficiently.

Environmental Law, the Economy, and Sustainable Development attempts to unravel the structural limitations of contemporary environmental law and chart a new regulatory landscape founded principally on market-friendly mechanisms. The contributors to the collection generally suggest that the key to improved environmental regulation lies 'in re-examining the bureaucratic structure of regulation itself' (p 298) than in the nature of environmental policy and social goals. What is different about this work is that it explores these regulatory issues in terms of the particular problems posed by environmental governance in 'multi-jurisdictional legal and political systems' (p 1). Although the authors confine themselves to analysing developments within the United States, the European Community and in international law, problems of environmental regulation in an inter-jurisdictional context obviously exist in many other federalist nations and regional contexts.

* School of Law, University of Warwick.

Bringing together views from fourteen scholars, the book canvasses a large array of contemporary regulatory problems that arise in multi-jurisdictional contexts. A number of the contributors have assimilated the virtues of recycling; several of the pieces were earlier published in similar or nearly identical form elsewhere. Prominent regulatory concerns include environmental interdependencies and externalities that transcend governmental borders, definition of the appropriate scale for allocating environmental management responsibilities, justifications for risk assessment, the distributional consequences of environmental regulation, the choice of environmental protection methods, and the scope for environmental constraints to international trade. An underlying theme is the relationship between environmental protection and economic policy, which manifests itself particularly in the trade liberalisation debates and the discussions about use of market mechanisms as a means of environmental policy. In their analysis of the excesses and shortcomings of current environmental regulation, the contributors lean to one side of the vast and complex debate. The authors 'do not view the welfare economic and non-economic dimensions of environmental policy as in fundamental conflict' and believe that 'well-crafted environmental regulatory institutions and instruments can advance both types of goals' (p 33). Most contributors believe these goals can be reconciled through decentralised institutional approaches harnessing market-based policy tools.

Despite the book's title, it actually says very little about 'sustainable development'. The latter concept has become enshrined in the European Community Treaty as a fundamental objective for the Community's institutions and appears extensively in national environmental legislation. There is no discussion relating to the policy content or regulatory requirements of sustainability in Richard Stewart's introductory chapter and only in the final chapter by Geoffrey Heal on 'Markets and Sustainability' does the book begin to address this topic seriously. Heal, is optimistic that markets are not intrinsically myopic or otherwise environmentally-unfriendly but can be 'reoriented' through appropriate 'legal infrastructure' (p 427). Such optimism ignores the insights by ecological economists such as Herman Daly and Robert Costanza, who have long argued that overall ecological limits and environmental quality objectives must be developed socially because, whilst the market may facilitate resource allocation, it lacks a mechanism governing scale – that is, the market has no intrinsic tendency to limit the scale of aggregate use of environmental resources within biosphere limits.

In today's world of trade liberalisation, free-flowing capital is said to gravitate towards those localities where the cost of meeting regulatory burdens is the lowest and profits are greatest. In a global regulatory market place, there is a risk that governments may ease domestic environmental regulations while seeking to out-bid one another to attract capital investment. Richard Revesz and Roger van den Bergh reject this 'race-to-the-bottom' thesis and favour devolving decision-making to lower level authorities. In a debate ignited by Revesz in the early 1990s, he argues that there is no race-to-the-bottom in state environmental standard-setting, and that inter-state competition can be welfare-enhancing rather than welfare-reducing. The argument has been criticised by Kirsten Engel, Daniel Esty and others. Revesz's thesis is grounded in neo-classical economic theory, which posits that ideally competitive markets lead to economically-efficient outcomes. Extending this theory to the public sector, Revesz reasons that competitive markets will likewise result in an efficient degree of local legislated environmental protection. The theoretical basis for the race-to-the-bottom argument, by contrast, is non-cooperative game theory, which emerged as a branch of economics largely because of the prevalent failure of the 'ideally competitive market' hypothesis in real-world economies. Inasmuch as governments

act strategically and the market for development-location rights is relatively imperfect, unbridled regulatory competition among jurisdictions will tend to foster inferior environmental standards. Revesz's rejection of central (federalist) environmental standards rests upon a 'presumption in favor of decentralization' (p 39) that may be rebutted if there is 'a systemic evil in letting states decide the level of environmental protection that will apply within their jurisdictions' (p 40). To Revesz, this 'presumption' is a substitute for empirical evidence regarding what is arguably, however, an empirical question. Only John Ferejohn, of all the contributors, strongly believes that decentralised regulation may be politically and economically inadequate to address inter-jurisdictional pollution spillovers, which, he argues, are usually better managed through higher level institutions.

Interestingly, when it comes to reconciling trade liberalisation rules and environmental policy, several contributors reject a decentralised approach whereby nations may unilaterally invoke process-based environmental trade restrictions to restrict imports. Fearing disguised trade restrictions and destabilisation of the global trading system, Frieder Roessler believes that the integrity of the GATT-WTO system should not be cross-linked to environmental policy concerns. Scott Barrett and Ernst-Ulrich Petersmann are prepared to make some concessions to the environmental agenda in trade rules, such as to address trans-jurisdictional pollution spillovers, but they do not question fundamentally the environmental sequelae of trade liberalisation systems.

One of the most contentious environmental issues facing regulators and communities today is management of genetically modified foods and organisms and hazardous substances. In their examination of environmental risk assessment and management processes, Stephen Breyer and Veerle Heyvaert portray a 'vicious circle' in which ill-informed community perceptions of risk, regulatory responses to those perceptions, and authorities' reliance on uncertain science generate problems of random and excessive regulation. Their solution is centralised and discretionary regulatory controls run by a cadre of experts. The advantage of such an institutional arrangement, assert Breyer and Heyvaert, is that it would harness the Weberian virtues of bureaucracy – rationalisation, expertise, insulation, and authority. However, diluting public input in the process of risk regulation may not remedy the quandaries posed by scientific uncertainty given the non-political sources of scientific uncertainty. Environmental issues are plagued by imperfect knowledge, as highlighted by the precautionary principle, one of the building-blocks of sustainable development policy. Surprisingly, however, there is no reference to this principle in the book's discussions of risk assessment. The precautionary principle is increasingly central to environmental risk assessment regulation in various countries (e.g., New Zealand's Environmental Risk Management Agency) and relevant international treaties (eg, the Biosafety Protocol, 2000). Sound risk assessment procedures are grounded not simply in coherent administrative decision systems but proceed from principled policy frameworks.

The contributors to the section on choice of environmental policy instruments tend to argue for a greater role for economic instruments (e.g., taxes, marketable permits and industry agreements) as an alternative or addition to traditional modes of command regulation, condemned for its inefficiencies, politicisation and limited success in actually reducing environmental problems. Richard Stewart, offering the most comprehensive essay, lucidly canvasses the pros and cons of economic instruments against command regulation, drawing principally upon lessons from the United States. Eckard Rehbinder, offering a European perspective, is more cautious in his assessment of economic instruments, noting their potential impact on

individual rights in environmental decision-making. Although Stewart also refers to distributional and other ethical issues, questions of cost efficiency and maintenance of market systems dominate the analyses of policy instruments. Stewart argues that we should eschew technology-based command regulations in favour of market-based incentives that harness the economic self-interest of each business in the service of environmental protection.

The discussion of choice of policy instruments tends to rehearse much that has already been said in this field, and does not consider the potentially more fruitful scope for regulatory reform concerning not the role of market instruments, but of *market organisations*. Market instruments for the environment such as pollution charges and tradeable rights can be suitable for reforming specific industries and sectors – they convey costs and benefits of resource use options, thereby facilitating allocative efficiency in resource use and technological innovation. But they tend to be less suitable for generating macro-economic structural changes derived from access to capital markets that shape the overall scale of aggregate resource use. Preliminary work by the European Commission and the United Nations Environment Programme has identified the role of financial organisations, such as banks, insurance companies and institutional investors, as potential surrogate environmental regulators through their corporate financing and pricing activities. These financial institutions may be in a position to restrict or condition financing to development projects that are environmentally questionable and to steer investment resources to environmentally sound sectors of the economy. The financial sector can thus be important for transmitting and indeed amplifying the effects of environmental concerns through the economy.

Overall, the value of this book is that it makes a case for a decentralised and economic-instrument based approach to environmental regulation as a solution to the challenge of efficiently co-ordinating environmental policy in a multi-jurisdictional setting. The focus on regional and multi-jurisdictional regulatory contexts provides an important bridge between scholarship that has focused on problems of national environmental law and analyses of international environmental law. Aside from the lack of engagement with the complexities of sustainable development policy, some of the institutional linkages and regulatory relationships necessary to promote environmental protection are brought into sharp relief by this work.

*Benjamin J. Richardson**

* School of Law, University of Manchester.

Information for Contributors

Contributions to and correspondence concerning the various sections of the Review should be sent to the relevant Editor:

Articles: Hugh Collins, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Legislation and Reports: R.C. Simpson, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Case Notes: Colin Scott, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Books for Review, Review Articles and Reviews: Neil Duxbury, Faculty of Law, University of Manchester, Oxford Road, Manchester M13 9PL.

General Correspondence: Tim Murphy, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Section Editors will be happy to discuss drafts and proposed contributions with contributors. All contributions should be typed, double spaced and include an estimate of the number of words, including footnotes. Articles should be preceded by a short abstract of approximately 100–150 words. Reports, Legislation and Case Notes should be preceded by a list of all reports, legislation and cases cited in the paper, displayed as keywords. Contributions should use footnotes which make cross-references within the text as sparingly as possible. Contributions must conform to the Review's house style; a style sheet is available on the MLR website or will be provided by the Production Editor on request.

Articles do not usually exceed 12,000 words (excluding footnotes). Longer or shorter articles can be considered

Subscription Information 2001

The Modern Law Review is published six times a year. Subscription rates are as follows:

Annual subscription for unbound parts only

UK and Europe	£108.00
Rest of the world	£114.00
The Americas	US\$186.00*
UK and European law teachers and law students	£25.00
Commonwealth law teachers and law students	£32.00/US\$50.00*

Annual subscriptions for Bound Volume Service: unbound parts plus a bound volume

UK and Europe	£159.00
Rest of the world	£169.00
The Americas	US\$266.00*
UK and European law teachers and students	£65.00
Commonwealth law teachers and law students	£68.00/US\$109.00*

Back Issues

Single issues of the unbound parts of the last three volumes are available from the publishers at the current price for unbound parts:

	£22.00 (£23.00 for ROW, US\$37.00 The Americas)*
Single bound volumes	£71.00 (£74.00 for ROW, US\$122.00 The Americas)*
Complete sets of bound volumes	£2071.00 (US\$3417.00 The Americas)*

*Canadian customers/residents please add 7% GST.

Orders, sample copy requests, claims and other subscription enquiries should be addressed to: Blackwell Publishers Journals, PO Box 805, 108 Cowley Road, Oxford OX4 1FH, UK. All subscriptions are supplied on a calendar year basis (January to December).

US Mailing

Periodicals postage paid at Rahway, New Jersey.

Postmaster: send address corrections to *The Modern Law Review*, c/o Mercury Airfreight International Ltd Inc., 365 Blair Road, Avenel, NJ 07001, USA (US mailing agent).

Microform

The journal is available on microfilm (16 mm or 35 mm) or 105 mm microfiche from the Serials Acquisitions Department, Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106, USA

VOL. 64, NO. 2

MARCH 2001

THE MODERN LAW REVIEW

PUBLISHED FOR THE MODERN LAW REVIEW LIMITED
BY BLACKWELL PUBLISHERS, OXFORD, UK AND BOSTON, USA

THE MODERN LAW REVIEW

Volume 64

EDITORIAL BOARD

J.S. Anderson
M.R. Chesterman
Lord Dahrendorf of Clare Market
Gerald Dworkin
J.M. Evans
Anthony Giddens
Cyril Glasser
J.A.G. Griffith

Carol Harlow
Harvey McGregor
T.M. Partington
I.G. Patel
Lord Wedderburn of Charlton, *Editor Emeritus*
Michael Zander

EDITORIAL COMMITTEE

Julia Black
Roger Brownsword
Hugh Collins, *Articles Editor*
W.R. Cornish
Neil Duxbury, *Reviews Editor*
Judith Freedman
Nicola Lacey
Martin Loughlin
Kate Malleson, *Production Editor*

W.T. Murphy, *General Editor*
R.W. Rawlings
S.A. Roberts
Colin Scott, *Case Note Editor*
R.C. Simpson, *Legislation and Reports Editor*
F.G. Snyder
Erika Szyszczak
Sarah Worthington, *Secretary*

Published for The Modern Law Review Limited by Blackwell Publishers.

© The Modern Law Review Limited 2001 ISSN 0026-7961

The Modern Law Review Ltd is incorporated in England as a company limited by guarantee under the Companies Act (Reg No 325 282) and is registered as a charity (Reg No 264072)

The Modern Law Review is published six times a year, in January, March, May, July, September and November by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main St, Malden, MA 02148, USA.

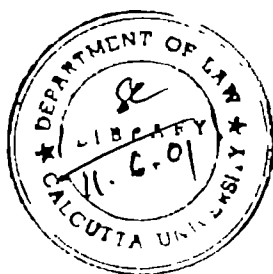
Information for contributors, subscription information and details about microfilm and advertising appear on the inside back cover.

The Modern Law Review website is at <http://www.lse.ac.uk/depts/law/mlr.htm>

Printed and bound in Great Britain by The Alden Press, Oxford
This journal is printed on acid-free paper.

COPYRIGHT: All rights reserved. Apart from fair dealing for the purposes of research or private study, or criticism or review, as permitted under the UK Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing of the Publisher, or in accordance with the terms of photocopying licences issued by organisations authorised by the Publisher to administer reprographic rights, such as the Copyright Licensing Agency (CLA) and the Copyright Clearance Center (CCC). Institutions with a paid subscription to this journal may make photocopies for teaching purposes free of charge provided such copies are not resold

CONTENTS



ARTICLES

- Unravelling *Osman* 159
Conor A. Gearty

- Risk and Remoteness of Damage in Negligence 191
Marc Stauch

- Culture, Autonomy and *Djulibinyamurr*: Individual and Community in the
 Construction of Rights to Traditional Designs 215
Kimberlee Weatherall

LEGISLATION

- New Labour's Welfare Reforms: Anything New? 243
Julian Fulbrook

REPORTS

- New Labour, Same Old Tory Housing Policy? 260
Dave Cowan and Alex Marsh

- The Criminal Cases Review Commission: Reporting Success? 280
Richard Nobles and David Schiff

CASES

- Forensic Immunity Post-*Osman* 300
Rosalind English

REVIEW ARTICLE

- Sexual Autonomy and Law 309
Mary Childs

REVIEWS

- Ibbetson: A Historical Introduction to the Law of Obligations 324
 Ripstein: Equality, Responsibility and the Law 326
 Rubio-Marín: Immigration as a Democratic Challenge: Citizenship
 and Inclusion in Germany and the United States 328

THE MODERN LAW REVIEW

Volume 64

March 2001

No 2

Unravelling *Osman*

Conor A. Gearty*

The judgment of the European Court of Human Rights in Osman v United Kingdom, decided in October 1998, has proved very controversial. Its implications for the UK law of negligence appear to be immense. Not the least of the complexities associated with the decision is the widespread perception that the reasoning of the Court is extremely difficult to understand, indeed, that it is at times contradictory. This article subjects the Osman judgment to close analysis. It explains the Court's approach to Article 6(1) by reference to the way in which the Strasbourg court has over many years developed its case law on this provision. The article includes an overview of the way in which, so far, the decision has been deployed in UK law. The author suggests that, whatever about the actual result of the case, the reasoning of the European Court in Osman is deeply flawed, and that the UK judiciary should be mindful of this fact when considering its deployment in domestic law.

By 28 October 1998, the Human Rights Bill had almost completed its passage through Parliament.¹ In a routine piece of business in the upper house on that day, a government minister is to be found referring to the then forthcoming fiftieth anniversary of the signing of the Universal Declaration on Human Rights as 'an important opportunity to re-affirm the universal nature of human rights',² and pointing to the Human Rights Bill as evidence of the administration's determination 'to give further effect to ECHR rights in domestic law so that people can enforce those rights in the United Kingdom courts.'³ That very day, the European Court of Human Rights in Strasbourg released its judgment in *Osman v United Kingdom*.⁴ The case involves two of the Convention rights (Articles 2 and 6) which are now part of UK law. Though not required to be followed, *Osman* will need to be dealt with in some way by British judges, both because the Human Rights Act itself so requires⁵ and because Strasbourg remains open to disappointed

* Professor of Human Rights Law, King's College London; founding member, Matrix Chambers. My thanks to K. D. Ewing, C. Harlow, M. Lunney and A. Tomkins who have read and commented upon earlier drafts of this article; I retain sole responsibility for the content, including any errors that there may be. My thanks are also due to the Leverhulme Trust for an award which has given me a much fuller opportunity than I would otherwise have had to conduct the research upon which this article is based, and to secure the time to be able to write up my conclusions.

1 The Bill received the Royal Assent on 9 November: HC Deb vol 319 col 32: 9 November 1998.

2 HL Deb vol 593 col 1910: 28 October 1998 (Baroness Symons of Vernham Dean, Under Secretary of State at the Foreign and Commonwealth Office).

3 *ibid* col 1912. The reference to the 'ECHR' is of course to the European Convention on Human Rights.

4 (1998) 29 EHRR 245; [1999] 1 FLR 193; [1999] *Family Law* 86.

5 Human Rights Act 1998, s 2.

domestic applicants even after full implementation of the measure. To put it at its mildest, the decision demonstrates that the 'universal nature of human rights' is of a more complex construction than that phrase would seem to suggest.

The basic facts of the case were simple. The applicants had sought to sue the Metropolitan Police Commissioner in the English courts for negligence, alleging a breach of a duty of care owed to them, which they claimed arose out of a culpable police failure to prevent a third party from fatally shooting a member of their family (the first applicant's husband, who was also the second applicant's father). The action was struck out in the Court of Appeal as disclosing no reasonable cause of action, on the basis that it was clear law after the House of Lords decision in *Hill v Chief Constable of West Yorkshire*⁶ that no action could lie against the police for negligence in a situation such as the one allegedly before the court, and that this was the case even if all the disputed facts were resolved in the claimants' favour.⁷ In its ruling, the Strasbourg court found that the police failure to protect the right to life of the applicants' husband and father did not amount to a breach of the qualified right to life to be found in Article 2 of the Convention.⁸ This aspect of the decision is a very important one with its own implications for official conduct, but it will not be further pursued here. Our interest lies exclusively with what the Strasbourg court had to say about Article 6.

The judges were unanimous that the procedure whereby the applicants' complaint of negligence came to be defeated in the domestic courts had infringed the first sentence of Article 6(1) of the Convention. At first glance, the terms of that sentence do not make this result as clear as the unanimity of the Court would suggest:

In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law.

Clearly what was involved here was the determination of a 'civil right' (in Convention terms this was the right to sue for damages for negligence), but equally clearly – one would have thought – the applicants had had a 'fair and public hearing'. Indeed they had had two such hearings, in the High Court and afterwards in the Court of Appeal, where the courts had, if anything, been 'over-fair' to them, in that all the disputed facts had been assumed to have been found in their favour. No question of any lack of openness or of delay having being alleged, how could this procedure have been found to have been in breach of this part of the Convention? The European Court's answer – that Article 6(1) was applicable because the applicants had asserted the right to sue with at least some prospect of success⁹ and that it was breached because the inflexible ruling in *Hill* meant that they were bound to fail¹⁰ – seems on a superficial reading at least to be somewhat contradictory. Also the European Court's invocation of the applicants' 'right to a court', upon which it based this part of its judgment,¹¹ is not easy to understand, given that no such right is expressly guaranteed in Article 6 and even if it were explicitly present, the applicants seemed (as earlier mentioned) to have had

6 [1989] AC 53.

7 *Osman v Ferguson and Another* [1993] 4 All ER 344. The application to strike out had been earlier dismissed in the High Court and so the matter had come before the Court of Appeal as an appeal against that decision by the Metropolitan Police Commissioner.

8 n 4 above, paras 101–123.

9 *ibid* paras 138–139.

10 *ibid* para 152.

11 See *ibid* para 147.

abundant opportunities to put their case before a succession of such tribunals. In any event, how could the right to a court have been good enough for them to have credibly asserted their right to sue but not good enough when it came to the substance of the court hearing itself?

Though still a relatively new decision, the Article 6 aspect of *Osman* has already come in for some very heavyweight criticism in the United Kingdom. Lord Browne-Wilkinson has observed that after *Osman*, 'extreme care ... must be taken in striking out claims in this confused and developing area of law.'¹² Lord Woolf has referred judicially to the fact that the 'possible consequences' of the case 'concern' him 'from a procedural point of view in the light of the new culture in litigation in this jurisdiction as a result of the introduction of the Civil Procedure Rules 1998 ...'.¹³ The United Kingdom judge on the European Court of Human Rights, Sir Nicolas Bratza, has referred extra-judicially to the decision, noting the fact that it has 'attracted so much hostile publicity in this country'.¹⁴ Lord Hoffman, also writing extra-judicially, has described the case as one which 'fills [him] with apprehension,' and which 'reinforce[s] the doubts [he has] had for a long time about the suitability, at least for this country, of having questions of human rights determined by an international tribunal made up of judges from many countries.'¹⁵ Tony Weir asserts in the *Cambridge Law Journal* that

[n]ations should decide for themselves whether public funds should be directed to victims of past malfunction in public services or used to reduce the number of such malfunctions in the future. ... In any case to answer this question in terms of 'human rights' is frankly absurd.¹⁶

Reflecting on the case in a recent decision in the House of Lords, Lord Browne-Wilkinson 'confess[ed] that [he found] the decision of the Strasbourg court extremely difficult to understand'¹⁷ and expressed 'the hope that the law applicable under Art 6 [would be] further interpreted.'¹⁸

Lord Browne-Wilkinson's aspiration is easier to state than to achieve. The opacity of the reasoning in *Osman* makes it particularly resistant to critical analysis. An added difficulty – indeed one of the central problems with *Osman* – is that it is only a recent manifestation of a line of Strasbourg cases in which similarly complex modes of legal reasoning have been displayed. The *Osman* decision does not come from nowhere; its various formulations have a pedigree in the Strasbourg jurisprudence, and a knowledge of these cases is essential to a proper understanding of the decision. But the matter is not extensively addressed in any of the many UK-based textbooks on human rights and the European Convention

12 *Barrett v Enfield Borough Council* [1999] 3 All ER 193, 197.

13 *Kent v Griffiths* [2000] 2 All ER 474, 484. There is more in a similar vein in *Palmer v Tees HA* [2000] PNLR 87 and *Aldrich v Norwich Union* (Court of Appeal, 30 July 1999). See further *S v Gloucestershire CC* [2000] 3 All ER 346.

14 Sir N. Bratza, 'The Implications of the Human Rights Act 1998 for Commercial Practice' [2000] EHRLR 1, 9.

15 'Human Rights and the House of Lords' (1999) 62 MLR 159, 164.

16 (1999) 58 CLJ 4, 7. M. Lunney, 'A Tort Lawyer's View of *Osman v United Kingdom*' [1999] KCLJ 238 is an exceptionally powerful critique. P. Craig and D. Fairgrieve, 'Barrett, Negligence and Discretionary Powers' [1999] PL 626 think the 'reasoning of the ECHR [to be] not convincing' (at 630). Also very critical is G. Monti, '*Osman v UK* – Transforming English Negligence Law into French Administrative Law?' (1999) 48 ICLQ 757. For an overview of the English law, written before *Osman* was decided, see J. Wright, 'Local Authorities, the Duty of Care and the European Convention on Human Rights' (1998) 18 OJLS 1. For a close analysis of the implications of the case for the police see P. Giliker, '*Osman* and Police Immunity in the English Law of Torts' (2000) 20 LS 372.

17 *Barrett v Enfield Borough Council* n 12 above, 198.

18 *Ibid* 200.

which have recently emerged.¹⁹ This is not surprising: the authors of such books have had huge ground to cover at a breakneck speed, since they have all sought to get into print before implementation of the Human Rights Act 1998 on 2 October 2000. More significantly, the precise issue central to *Osman* – the substantive right of access to a court – is not dealt with in the standard texts on the Convention, one of which was published as recently as 1998.²⁰ The line of cases behind *Osman* may be long and (now) venerable but the diversion down the particular route it has followed is of fairly recent origin.

The only way to understand *Osman* is to go back to first principles, and to trace how the relatively simple guarantees of Article 6(1) have been transformed into the apparently contradictory and over-complex reasoning that has been the focus of so much criticism. The story is necessarily a long, and at times circuitous adventure-ride through the jurisprudence of the Strasbourg court. Having first analysed the early case law, this article will then proceed to an examination of the various decisions of the Court from which *Osman*'s version of the right to a court has been deduced. In doing this we will show what the pressures have been which have wrenched the Strasbourg case law in such a succession of what will be seen to be peculiar directions. With this ground prepared, the article will then critically analyse the reasoning in the *Osman* decision itself. The effect of our inquiry will be to confirm and reinforce the doubts to which the decision has given rise. Having come to this critical conclusion, we will then consider the reception of *Osman* and the *Osman* line of cases into UK law, and make certain suggestions as to the right approach to be adopted now that implementation of the Human Rights Act has finally thrust the issue squarely upon the courts.

The concept of a 'determination' of a civil right

As noted above, the relevant part of the first sentence in Article 6(1) is in the following terms:

In the determination of his civil rights and obligations ... everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law.

The latter part of this sentence is reasonably clear. The entitlements there set out are the kind of procedural guarantees that are to be found in many legal systems. The key question begged by the Convention's formulation relates to when these safeguards can be said to be required; in other words when are 'civil rights' the subject of a 'determination' for the purpose of attracting Article 6(1) protection? The easy answer, which is right but not complete, is that Article 6(1) applies to all ordinary civil actions, in which 'civil rights' (asserted by claimants) are routinely 'determined' by an adjudicating court. The Convention is certainly a guarantor of due process in this arena of civil litigation, and the European Court of Human

19 Of the books that do mention *Osman*, see A. Lester and D. Pannick, *Human Rights Law and Practice* (London: Butterworths, 1999) para 4.6.5; J. Wadham and H. Mountfield, *Blackstone's Guide to the Human Rights Act 1998* (London: Blackstone, 1999) 89–90; J. Coppel, *The Human Rights Act 1998. Enforcing the European Convention in the Domestic Courts* (Chichester: John Wiley, 1999) 243. For a somewhat fuller treatment see S. Gross, J. Beatson and P. Duffy, *Human Rights. The 1998 Act and the European Convention* (London: Sweet and Maxwell, 2000) 225, 233–234.

20 In P. van Dijk and G. J. H. van Hoof, *Theory and Practice of the European Convention on Human Rights* (The Hague: Kluwer, 3rd ed, 1998) there is no hint of such a substantive right with the question of access to a court being dealt with as a quite separate issue within Art 6(1): see 418–428.

Rights has frequently held that this is the case even where the state itself is involved as a party. In the early days of the Court's jurisprudence, these cases never raised 'an access to court' issue such as was relied upon in *Osman* but rather were processed under one or other of the substantive heads in Article 6(1) itself. Indeed this is still occasionally the case today; in *Kerojärvi v Finland*²¹ for example, a refusal by the Finnish Supreme Court to make documents available to the applicant in a compensation case was held to infringe the fairness requirement of Article 6(1), with there being no question of any argument based on access to the court being relied upon. In *Mialhe v France (No. 2)*²² where the issue concerned the denial of access to documents held by the revenue authorities during the administration stage of the Tax Offences Board's deliberations on the applicant's tax affairs, the Court treated the point as involving solely an issue of 'equality of arms' under the substance of Article 6(1). So why was *Osman* not just an ordinary Article 6(1) case, with the issue of the strike-out procedure being tested against the substantive safeguards appearing in the text of the Article?

To answer this question, we need to take a short detour into the foundational jurisprudence of Article 6(1), for without the early ambitions that many Strasbourg judges had for the provision (and which they sought to impose on it) no *Osman*-style issue would ever have arisen. Had Article 6(1) been left merely as a cover for defective civil litigation, it would have been clear and easy to grasp. The only question in *Osman*, for example, would have been whether the applicants had had a 'fair and public hearing'. Left only to such cases, however, Article 6(1) would have been relatively unimportant, since the civil action is an area in which member states of the Council of Europe have generally long committed themselves to just the kind of procedural rights that are required by Article 6(1). So this part of the Article would have ended up being largely duplicatory of domestic law and would therefore have generated little of its own jurisprudence, apart that is from cases on delay and on the kind of fairness issue that should have been to the fore in *Osman*. And it would not have extended to administrative law at all, a subject which is covered explicitly in no other Convention Article. That there was more to the civil aspect of Article 6(1) than met the eye, and that the administrative law lacuna was not to be wholly ignored, first became evident as early as the thirteenth case before the Court, *Ringeisen v Austria*.²³

In this case the Court extended its remit to decisions by the Austrian Regional Real Property Transactions Commission in relation to the approval of certain contracts for the sale of land. The Strasbourg court applied the procedural requirements of Article 6(1) to the Austrian tribunal, notwithstanding its technically non-judicial status. In an important ruling, the judges unanimously declared that Article 6(1) was wide enough to cover 'all proceedings the result of which [were] decisive for private rights and obligations... Although it was applying rules of administrative law, the Regional Commission's decision was to be decisive for the relations in civil law ("*de caractère civil*") between Ringeisen and the other party to the contract.'²⁴ This was 'enough to make it necessary for the Court to decide whether or not the proceedings in this case complied with the requirements

21 European Court of Human Rights, 19 July 1995. See, further, *Fouquet v France* (1996) 22 EHRR 279, *Mantovanelli v France* (1997) 24 EHRR 370.

22 (1996) 23 EHRR 491.

23 (1971) 1 EHRR 455. Note that in the citations from the EHRR in this article I have denoted the year in which a European Court of Human Rights case has been decided as well as the volume number of the European Human Rights Reports in which the decision is to be found.

24 *ibid* para 94.

of Article 6(1) of the Convention.²⁵ *Ringeisen* made it clear very early in the Strasbourg court's life that Article 6(1) could reach past ordinary, *inter partes* civil actions and clasp on to administrative decisions taken by state bodies of a quasi-judicial sort.²⁶ In both ordinary judicial proceedings and *Ringeisen*-type tribunal adjudications involving 'civil rights' (in the Article 6(1) sense), it will be obvious that there is the requisite 'determination' of the civil right in issue for Article 6(1) purposes.²⁷ There are proceedings under way. Battle is joined between two parties and the outcome is decisive for the asserted right, one way or the other. This approach has survived in the case law, subject only to a sensible gloss on remoteness which was first postulated in the later case of *Le Compte, Van Leuven and De Meyere v Belgium*²⁸ to the effect that 'the result of [any such] proceedings [had to] be directly decisive for the civil right', with the Court considering that 'a tenuous connection or remote consequences [would] not suffice for Article 6 (1)...'.²⁹ It accounts for many of the cases on Article 6(1) which are relatively straightforward despite arising out of disputes which are not strictly civil in nature.

One point raised by *Ringeisen* was what was meant by a 'civil right'. This has been an issue of great concern since that decision and a substantial body of case law has grown up around this very question.³⁰ This is, however, not the topic of this article. The relevant question begged by *Ringeisen* for our purposes was whether Article 6(1) went further than the sort of proceedings that had arisen in that case. To resolve this issue, the European Court needed to confront what was meant by a 'determination' of a civil right for Article 6(1) purposes. The word was immensely problematic, probably more so than any other term in the whole Convention. Did this idea of a 'determination' necessarily involve a traditional (or a *Ringeisen*-style) legal dispute? Or was it a much broader and less formal concept, involving merely an interference (albeit a serious one) with the right in issue? At first glance, the English word 'determination' is more suggestive of the second than the first of these meanings. To 'determine' something is to bring it to an end or (perhaps) otherwise drastically to limit it in scope. This suggests a wide meaning, embracing any serious interference with, or cessation or extinction of a right, regardless of whether this takes place in a judicial, quasi-judicial or other kind of formally disputatious circumstance. If the English version of Article 6(1) were its only official construction, then Article 6(1) would be potentially very broad indeed, with the procedural guarantees in the Article bursting out of their civil context to embrace all officially-inspired 'determinations' of 'civil rights' wherever they occurred (the police station; the local authority committee room; the planning officer's desk, the central government ministry).

²⁵ *ibid.*

²⁶ For further discussion of the origins of Art 6(1), see the relevant sections in the collected edition of the *Travaux Préparatoires* (eight volumes) published by Martinus Nijhoff between 1975–85. A very useful summary of the background to Art 6(1) is to be found in the dissenting opinion of Judges Ryssdal, Bindaschelder-Robert, Lagergren, Matscher, Sir Vincent Evans, Bernhardt and Gering in *Feldbrugge v Netherlands* (1986) 8 EHRR 425, 444–445.

²⁷ See following the *Ringeisen* precedent: *Sramek v Austria* (1984) 7 EHRR 351; *Ertl v Austria* (1987) 10 EHRR 255; *Erkner and Hofauer v Austria* (1987) 9 EHRR 464; *Polas v Austria* (1987) 10 EHRR 231; *Hakansson and Sturesson v Sweden* (1990) 13 EHRR 1; *Stallinger and Kuro v Austria* (1997) 26 EHRR 81.

²⁸ (1981) 4 EHRR 1.

²⁹ *ibid* para 47. The case-law on causation is well-treated in van Dijk and van Hoof, n 20 above, 397–398. For a recent decision on the point see *Balmer-Schafroth v Sweden* (1997) 25 EHRR 598.

³⁰ There is a good survey, now a little out of date, in van Dijk and van Hoof, n 20 above, 398–406.

The French equivalent of 'determination' however points in a much more limiting direction. According to this alternative official version of the Convention, the safeguards of Article 6(1) extended only to '*contestations sur (des) droits et obligations de caractère civil*'. The idea of '*contestation sur*' is far more suggestive of a legal dispute than is the English language equivalent. Whereas the English concentrates on the substantive point – what has happened to the right? – the French, with its emphasis on '*contestations*' or claims, seems more concerned with the question, where has it happened? The point could be dodged in the ordinary civil cases before the Court on Article 6(1) points and even in *Ringeisen* because there happened to be in all of these cases an official decision with respect to civil rights which was both a 'determination' in the English sense and the consequence of a formal dispute in the strongest sense of the French idea of '*contestation*'. But it could not be avoided for long, particularly when applicants began to try to push the boundaries of Article 6(1) ever further back – way beyond the ordinary civil suit – in order to be able to get access to the procedural guarantees there on offer.

A path-breaking early case on this point was *Le Compte, Van Leuven and De Meyere v Belgium*³¹ in which Article 6(1) was applied to the disciplinary proceedings of the Belgian medical association. To the respondent state's argument that there was no "*contestation*" (dispute), in the sense of two conflicting claims or applications' on the facts in issue in the case,³² the Court replied that '[c]onformity with the spirit of the Convention requires that this word should not be construed too technically and that it should be given a substantive rather than a formal meaning besides, it has no counterpart in the English text of Article 6(1)'.³³ Then, seeming to resile from such a general approach and in an early example of the uncertainty that was to pervade the case-law in this area, the Court sought to split the difference between the two positions by declaring that '[e]ven if the use of the French word "*contestation*" implies the existence of a disagreement, the evidence clearly shows that there was one in this case.'³⁴ The applicants had opposed the charges made against them before their professional body and had then appealed against the decisions in a domestic process that had gone through a number of appellate levels before ending in Strasbourg. So there had clearly been an argument or a dispute which had reached levels of formality even if it was not precisely analogous with a civil claim (or even a *Ringeisen*-type dispute).

After *Le Compte*, therefore, the 'determination' of a 'civil right' for Article 6(1) purposes seemed neither to embrace any substantial interference or extinction of the right without more (a simple 'determination') nor to extend only to these same end-results when mediated through formal legal proceedings (a 'dispute' in a judicial or quasi-judicial setting). The phrase was neither wholly substantive (the English version) nor wholly procedural (the French gloss). Instead what the Court seemed to be saying was that what was essential was some kind of legal framework within which a disagreement about the state's action *vis-à-vis* a civil right could be ventilated.³⁵ Once this framework could be identified, Article 6(1) protection took

³¹ n 28 above.

³² *Ibid* para 45 of the judgment of the Court, citing counsel for the government.

³³ *Ibid*. The Court noted that the word 'dispute' did however appear in Art 49 of the Convention.

³⁴ *Ibid* para 45.

³⁵ cf *Moreira de Azevedo v Portugal* (1990) 13 EHRR 721, para 66: 'In so far as the French word *contestation* would appear to require the existence of a dispute, if indeed it does so at all, the facts of the case show that there was one.'

effect, and indeed the *Le Compte* decision itself went on to set out clear guidelines on how the safeguards in the Article could be met and at what level in the legal process.³⁶ It remains an important authority on these points to this day.

The arguable and unarguable case

Having muddied the waters on whether a dispute was actually required and if so what sort of dispute, *Le Compte* inevitably left open the issue of how one could know when a disagreement had reached the requisite degree of seriousness to amount to a 'dispute about'/'determination of' a right for Article 6(1) purposes. Obviously this would be as clear as day if the victim of a determination had had a full judicial process available to challenge the action, or had had access at least to a tribunal along *Ringeisen* lines. *Le Compte* also demonstrated that a dispute between a person and his or her professional body which engaged a formal administrative procedure also fitted into Article 6(1), but none of these cases provided very much more than this level of guidance. An even larger question was begged by *Le Compte*: what would be the situation where state law was so clear in relation to a civil right that it could not be said that the interference with it was subject to any meaningful dispute or disagreement for the purpose of unlocking Article 6(1) protection? In the very early case of *Golder v United Kingdom*,³⁷ the Court had suggested that Article 6(1) secured 'to everyone the right to have any claim relating to his civil rights and obligations brought before a court or tribunal.'³⁸ In *Le Compte*, the Court explained *Golder* further by saying that it stood for the proposition that 'Article 6(1) may ... be relied on by anyone who considers that an interference with the exercise of one of his (civil) rights is unlawful and complains that he has not had the possibility of submitting that claim to a tribunal meeting the requirements of Article 6(1).'³⁹

This emphasis on a subjective perception of domestic unlawfulness would be one way of determining which disagreements amounted to 'determinations' of civil rights for Article 6(1) purposes and which did not. Applicants who could demonstrate their belief that their 'civil rights' (as defined in Article 6(1)) were being unlawfully interfered with would have access to Article 6(1)'s substantive safeguards regardless of what kind of national body it was that was doing the interfering. But this approach did not remove the risk that states could skewer Article 6(1) to their advantage by taking unequivocal action effectively to state the law in a way which simply left no room for any assertion of this sort by an applicant. If this were to happen then the independent meaning which the Court had repeatedly emphasised it had the power to give to the term 'civil rights' in Article 6(1) would be vulnerable to being rendered simply irrelevant by national action which frontally and unequivocally assaulted such rights. Were this to happen, Article 6(1) would have been turned on its head: the more a state left its conduct open to challenge, the more Article 6(1) applied; the more it unequivocally

36 See in particular *Le Compte, Van Leuven and De Meyere*, n 28 above, paras 51(a) and 51(b) dealing with when in administrative law a public authority needs to be regarded as bound by Art 6(1) and when decisions of public authorities which are flawed from an Art 6(1) perspective can be redeemed by an appellate process meeting the demands of Art 6(1). See now *Kingsley v United Kingdom* European Court of Human Rights, 7 November 2000.

37 (1975) 1 EHRR 524.

38 *ibid* para 36.

39 *Le Compte, Van Leuven and De Meyere v Belgium*, n 28 above, para 44. To similar effect is *Sporrong and Lönnroth v Sweden* (1982) 5 EHRR 35, paras 79–80.

and unqualifiedly immunised its action from review, the less of a 'dispute' about alleged unlawfulness did Article 6(1) have upon which to bite. The point nearly arose in *Sporrong and Lönnroth v Sweden*⁴⁰ where the *Le Compte* test of domestic unlawfulness was repeated with the apparent addition of a requirement that the perceived dispute be 'serious' in nature.⁴¹ Only some very neat – and controversial – footwork by the Court in that case managed to conjure up the necessary 'dispute' from apparently very clear law interfering with a civil right in a way which seemed – until the majority of the Court looked very closely – to harbour no possibility of legitimate challenge.⁴² But the point could be no longer avoided in 1985, when *Ashingdane v United Kingdom*⁴³ came up for decision.

In this case, the applicant found his civil action against a government department and a local authority was doomed to defeat on account of the application to him of section 141 of the Mental Health Act 1959:

(1) No person shall be liable ... to any civil ... proceedings to which he would have been liable apart from this section in respect of any act purporting to be done in pursuance of this Act ... unless the act was done in bad faith or without reasonable care.

(2) No civil ... proceedings shall be brought against any person in any court in respect of any such act without the leave of the High Court, and the High Court shall not give leave under this section unless satisfied that there is substantial ground for the contention that the person to be proceeded against has acted in bad faith or without reasonable care.

Litigation such as *Ashingdane* desired to initiate was not prohibited, nor was it practically restricted in the sense that he would be physically prevented from filing his claim; rather his action was made the subject of a set of drastic, statute-imposed conditions which made victory in his case impossible: he could not point even to an arguable case of bad faith or lack of reasonable care on his particular set of facts.

When the case came before it, the European Court of Human Rights did not conclusively determine which 'civil right' was in issue since it had 'come to the conclusion that, even assuming Article 6(1) to be applicable, the requirements of Article 6(1) [had not been] violated.'⁴⁴ On the point of central concern for this Article, the question of whether this (assumed) civil right had been 'determined', the Court began by citing *Golder* as having established the right of access to a court and by then going on to describe this right as part of a broader 'right to a court' guaranteed by Article 6(1). This in turn was a general right which – echoing but not replicating *Le Compte*⁴⁵ and *Sporrong*⁴⁶ – could 'be relied on by anyone who considers on arguable grounds that an interference with the exercise of his (civil) rights is unlawful and complains that he has not had the possibility of submitting that claim to a tribunal meeting the requirements of Article 6(1).'⁴⁷ Though as we have seen the notion of 'unlawfulness' had cropped up before, the

40 *ibid.* To similar effect is *Tre Traktorer Aktiebolag v Sweden* (1989) 13 EHRR 309. An important early authority from the Commission on the point is *Kaplan v United Kingdom* (1980) 4 EHRR 64, paras 163–4 of the Commission Report.

41 n 39 above, para 81.

42 Other cases from this period which worked hard to orchestrate a dispute for Art 6(1) purposes include: *Bentham v Netherlands* (1985) 8 EHRR 1; *Van Marle v Netherlands* (1986) 8 EHRR 483; *Bodén v Sweden* (1987) 10 EHRR 367; *Pudas v Sweden* (1987) 10 EHRR 380.

43 (1985) 7 EHRR 528.

44 *ibid* para 54.

45 n 28 above.

46 n 39 above.

47 *Ashingdane v United Kingdom*, n 43 above, para 55, explicitly citing both *Le Compte*, *Van Leuven and De Meyere v Belgium*, n 28 above, and *Sporrong and Lönnroth v Sweden*, n 39 above.

idea of a threshold requirement of 'arguability' was a new addition to the case-law. It clearly constituted a higher obstacle to the application of Article 6(1) than the subjective tests of *Golder* and *Le Compte*, though it could perhaps be said to have been an extension of *Sporrong and Lönnroth's* apparent insistence on the seriousness of any dispute before Article 6(1) could be applied to it.⁴⁸ In all probability, the test emerged as one rooted in arguability because the Court was at this time developing exactly just such a threshold for the application of Article 13, and that provision's right to an effective remedy for the violation of Convention rights had many superficial similarities with Article 6(1).⁴⁹

Whatever its origin, applying this threshold test (as we shall henceforth call it) to the facts before the Court meant that Ashingdane was bound to fail as he was quite unable to establish any arguable case of any unlawful action against him. Indeed it was precisely his inability to supply arguable grounds for his various legal assertions which had led to his speedy and unequivocal defeat in the English courts. Ashingdane failed to establish an arguable case, of course, not because he was cantankerous or over-optimistic or a vexatious litigant, nor because he was (*Golder*-like) physically (and probably unlawfully) denied access to the courts, but because the law had been clearly designed to debilitate an action such as his. Section 141 of the 1959 Act had loaded the dice against him, even though it continued to be thrown by a croupier who was transparently honest and fair in the way he (or she) managed the game. The vital importance of the *Ashingdane* case from the point of view of principle is that, following the lead of the Commission,⁵⁰ the Court grasped the dangerous implications of its threshold test for European oversight of the operation of Article 6(1).

The Court noted first that the applicant had had access both to the High Court and the Court of Appeal where he had discovered that 'his actions were barred by operation of law' and that because of this it could be said that the applicant 'had [had] access to the remedies within the domestic system', even if it was only to find that he had no arguable case.⁵¹ But, instead of drawing the conclusion earlier prepared for, that therefore Article 6 was not applicable, the Court went on to say that this fruitless recourse to the courts did 'not necessarily exhaust the requirements of Article 6(1)'. The arguable case of domestic unlawfulness was not to be left as the only gateway to Article 6(1) protection, but was to be joined by a new test of compliance with Article 6(1), to apply only in situations where this threshold could not be passed by applicants. In a key passage the Court stated that even where the arguability threshold was not met,

[i]t must still be established that the degree of access afforded under the national legislation was sufficient to secure the individual's 'right to a court', having regard to the rule of law in a democratic society.⁵²

In other words, the Court reserved to itself the final decision as to whether the legally-sanctioned deprivation or restriction of access to the courts in any given case was such that it should in itself be deemed an infringement of Article 6(1). We shall henceforth call this the 'European fallback test'. Though in view of the way

48 See text at n 41 above.

49 A point made in the concurring opinion of Judge Martens (approved by Judges Foighel and Pekkanen) in *Salerno v Italy* (Series A No 245-D, 12 October 1992). On Article 13 see in particular *Silver v United Kingdom* (1983) 5 EHRR 347, para 113. See further B.M. Zupanec, 'Access to Court as a Human Right according to the European Convention on Human Rights' (2000) 9(2) Nott LJ 1.

50 See the Report of the Commission in *Ashingdane v United Kingdom* Series B, vol 76, at paras 91-92.

51 n 43 above, para 56.

52 *ibid*, para 57.

the case law has developed, one would have wanted the Court to have been even more explicit than it was, it was clear enough from the Court's reasoning that this test could only apply where the threshold test had not been passed and that, once it was brought into play, it would take effect notwithstanding that the impugned deprivation or restriction of access had been clearly set out in national law, and in respect of which therefore no issue of unlawfulness could have arisen at the domestic level (and thus by definition no arguable case): indeed this was the whole point of the test. On the Court's view of Article 6(1), therefore, the threshold to Article 6(1) proper was to be the arguability test rooted in national law, but where this could not be successfully negotiated a new doorway appeared before the applicant, through which underlying European supremacy could continue to assert itself *vis-à-vis* national regimes judged inappropriate in their approach to access. This was not a doorway to Article 6(1) proper however, but rather to a new, subsidiary test of access. To this limited extent, it is right to say that Article 6(1) as interpreted in *Ashingdane* does indeed involve a new substantive right of access to a court.⁵³

Having opened this new door to a different and more basic kind of Article 6(1) protection, the Court then applied it in *Ashingdane*, engaging in a careful and frequently cited examination of the principles underpinning the right of access to a court which had been implied by *Golder* into the provision. The details of this analysis need not detain us here, involving as it does the application of well-known principles of legitimacy and proportionality,⁵⁴ except to note that it produced a positive result for the respondent government, with the Court holding that, on the facts before it, the impugned legislation involved a justifiable restriction on the applicant's right to sue. Had it held otherwise, of course, the effect would have been to have held that, at least on the facts before it, the explicit decision by the UK Parliament to restrict litigation on the determination of a particular 'civil' right in the field of mental health had amounted to an infringement of Article 6(1). Such a ruling could easily have proved controversial, since it would have involved the Court in tackling directly the judgment of a democratically-elected legislature in a country still committed to parliamentary sovereignty. (Indeed it could be said that the perceived need for this fallback European test arose because of the UK's commitment to parliamentary sovereignty, since it was precisely the inability of the domestic courts in this jurisdiction to query legislation which meant that the arguable test of unlawfulness of the type developed in *Ashingdane* (the threshold test) would be so frequently not met in Britain. This was in contrast to most other Council of Europe member states where credible arguments of unlawfulness could be much more easily raised against legislative acts, by reference for example to a local written constitution or a domestic bill of rights.)

The 'right to a court' and the crisis of parliamentary sovereignty

Within a year of *Ashingdane*, the European Court had to confront the same issue as in that case in far more sensitive political circumstances. Perhaps because it was destabilised by the highly sensitive subject matter, this was a decision in which the Court was to make the first of the wrong turnings which have dogged so much of

⁵³ My thanks to Mark Lunney for making this point.

⁵⁴ See in particular *Ashingdane v United Kingdom*, n 43 above, para 57.

its Article 6(1) case law. *James v United Kingdom*⁵⁵ arose out of legislation introduced by the UK government of Harold Wilson⁵⁶ to give certain leaseholders the right in certain clearly defined circumstances to purchase the freehold of the property in which they resided. The legislation provided that disputes about the entitlement of a leaseholder to enfranchisement were to be referred to the county court, while disagreements about the level of the price to be paid to the landlord were to be settled by reference to a newly established Leasehold Valuation Tribunal. The latter tribunal was given a number of statutory criteria by reference to which it was required to come to its decision. Many freeholders were naturally unhappy about what they saw as the compulsory acquisition of their property, and a number of those involved in managing the Westminster estate (a likely significant loser under the new regime) challenged these legislative arrangements before both the European Commission and, subsequently, the European Court.

The main issue concerned the scheme's compatibility with the property guarantee in the first protocol, and as with *Ashingdane* the Article 6(1) point was slightly tagged along at the end. Still the applicants seemed to have a strong argument. Here was an undoubted civil right, the right to property.⁵⁷ It had been determined by Parliament in a most unequivocal fashion. True there was a right of access to the county court to challenge the enfranchisement of any particular tenant, but any such adjudication could only take place within the parameters set by the enfranchisement legislation, and this had clearly set out an absolute right to such an enfranchisement if certain objective and easily discernible conditions were met. There was, as the applicants put it, 'no question of individual merits, no question of hardship' which could be 'made susceptible of review by any court or tribunal'. Accordingly they contended that this framework of law amounted to a breach of Article 6(1). The threshold test in *Ashingdane* could not, however, be surmounted, since precisely because of the clarity of the legislation, no arguable case could be made as to the unlawfulness of the expropriation (unless of course the statutory conditions were not met, something that was not being alleged in this case). Like the court in *Ashingdane*, the county court could be guaranteed to act with scrupulous fairness, but it was equally certain, on account of the nature of the law that it had to apply, that the applicants would lose.

Faced with the insurmountability of the arguability threshold test in *Ashingdane*, it might perhaps be supposed that in oral and written argument, counsel for the applicants would have invited the Court to apply the fallback European test in *Ashingdane*, and in doing so therefore have asked the Strasbourg judges to consider whether 'the degree of access afforded under the national legislation was sufficient to secure the individual's "right to a court", having regard to the rule of law in a democratic society.'⁵⁸ It is perfectly possible that a legitimate case could have been made that an elected legislature is entitled to hamstring access to a court in a *James*-style way, particularly when engaged in the execution of a social policy which has been the subject of a manifesto pledge, much official consultation, and a thorough debate in both legislative chambers. An analysis of this sort might even have enabled the Court to have made some important statements of principle about the connection between the needs of 'a democratic society' and the potentially

55 (1986) 8 EHRR 123. Compare *Klass and Others v Federal Republic of Germany* (1978) 2 EHRR 214 where the issue nearly arose in the context of the applicants' contention that the German Basic Law infringed the Convention.

56 Leasehold Reform Act 1967 (as amended).

57 A point made clear in *Sporrong and Lönnroth v Sweden*, n 39 above.

58 As the Court put it in that case: see *Ashingdane v United Kingdom*, n 43 above, para 57.

immobilising effect of too obsessive a commitment to 'the rule of law'. It would probably have led to defeat for the applicants on this second European question as well as on the arguability threshold point, as it had done in *Ashingdane*, with the result that – as in *Ashingdane* – the UK government would have survived the Article 6(1) challenge.

However, this was not the approach the Court chose to adopt. The issue of Article 6(1) was not extensively canvassed before the Court.⁵⁹ Its treatment of Article 6(1) occupies four paragraphs of its judgment, and only one of these deals with the substantive law. *Ashingdane* is nowhere referred to in this part of the decision. Its two-part test is not mentioned, though parts of it are borrowed in the new resolution of the issue, which was set out in the following terms:

Article 6(1) extends only to 'contestations' (disputes) over (civil) 'rights and obligations' which can be said, at least on arguable grounds, to be recognised under domestic law: it does not in itself guarantee any particular content for (civil) 'rights and obligations' in the substantive law of the Contracting States.⁶⁰

The notion of 'recognition' here seems to have replaced the idea of 'unlawfulness' which had appeared in *Ashingdane*'s threshold test. Perhaps 'unlawfulness' was too crude a reminder of the surrender that is involved in allowing the member states to set the pace, but that they were now (*contra Ashingdane*) being permitted to do so is clear from the last clause in the above sentence. In this particular context, 'recognition' seems merely to be a positive version of unlawful, with both producing the same result: a right the interference with which can be challenged as unlawful is one that is recognised, whereas a 'civil right' that is taken with no recourse to the law, or a recourse to the law that is bound to fail, is one that is not 'recognised' by the state. While merely driving this point home, the final line quoted above, to the effect that Article 6(1) does not guarantee 'any particular content ... in the substantive law' of Member States, was particularly unfortunate because of the distinction it was later inevitably thought to suggest between procedural and substantive rights, a distinction without a difference in the context of Article 6(1) but one which, as we shall see, was later to return to contaminate the Strasbourg case-law with further unnecessary confusion.⁶¹

To confirm and consolidate the shift towards national relativity presaged in *James* by the Court's reworking of the *Ashingdane* threshold test, the second limb of the *Ashingdane* formulation – the fallback European test – was simply jettisoned without explanation. 'In so far as the applicants may have considered that there was cause for alleging non-compliance with the leasehold reform legislation, they had unimpeded access to a tribunal competent to determine any such issue.'⁶² The quality of that access was no longer an issue. No subsidiary test asserting a long-stop European supervision any longer applied. Entry into Article 6(1) was determined by domestic 'recognition' of the civil right and by nothing more.

That the Court was trying to cope with the question of parliamentary sovereignty cannot be in any doubt. It sought support for its analysis in a clear statement to this effect:

⁵⁹ It did not figure at all in argument before the Commission, being raised for the first time in the memorial of the applicants filed with the registry of the Court: see *James v United Kingdom*, Series B, vol 81, 72 and 113.

⁶⁰ *James v United Kingdom* (1986) 8 EHRR 123, para 81.

⁶¹ See below text at n 83.

⁶² *James v United Kingdom*, n 60 above, para 81.

Article 6(1) does not require that there be a national court with competence to invalidate or override national law. In the present case, the immediate consequence of the British legislation in issue is that the landlord cannot challenge the tenant's entitlement to acquire the property compulsorily in so far as the acquisition is in conformity with the legislation.⁶³

We should note in passing the particular nature of this problem. In jurisdictions with written constitutions, it is only those documents that can obliterate arguable cases of unlawfulness in the way that Acts of Parliament can in the UK, since all subsidiary law (including legislation) can be challenged by reference to them. There being no written constitution in the UK, far fewer laws are theoretically subject to dispute in this way. The difficulty lies not with the underlying purpose of the analysis but with its execution. As suggested above, the Court could have deployed the *Ashingdane* formula to achieve a far more elegant reconciliation between due process and democracy while also preserving intellectual coherence. Instead the effect of the formula produced in *James* was to add a fresh layer of confusion to Article 6(1). Viewed in isolation from its democratic context, and apparently subversive of the whole trend towards autonomous definition in the rest of Article 6, it is a dictum from which later decisions have sought to resile, in ways that have added their own layers of confusion to an already hugely complicated picture, producing among other problems the decision in *Osman*.

Life after *James*: subservience to national law?

The *James* 'arguable grounds for recognition' formula took root very quickly.⁶⁴ A particularly needless development was its speedy movement sideways into the mainstream of Article 6(1) jurisprudence, so that even cases raising no issue whatsoever of any sort of access to a court were soon also reiterating the test as a preliminary requirement before the Article 6(1) safeguards could be guaranteed.⁶⁵ Adoption of the test marked the extinction of the right of access line of cases as it had up to then been understood, with applicants who failed the preliminary threshold test now finding themselves shut out of Article 6(1) protection without any opportunity of being able to rely on *Ashingdane*'s fallback European test. The *James* approach was unequivocally confirmed in *Lithgow v United Kingdom*,⁶⁶ decided just five months later. In *Lithgow*, the Court repeated the key section from *James* on Article 6(1) word for word.⁶⁷ Though that case also involved the enactment of legislation providing for the taking of property, this time in the form of taking various companies into public ownership, the only issue directly addressed by the Court under Article 6(1) related to the civil right of the applicants to compensation under the scheme set up by the Act. Article 6(1) was said to be applicable 'in so far as the applicants may reasonably have considered that there

⁶³ *ibid*

⁶⁴ Note however that in the important and frequently cited case of *Editions Periscope v France* (1992) 14 EHRR 597, the phrase 'sufficiently tenable' rather than 'arguable grounds' appears, but this formula has not taken root in the case-law in the way that other aspects of this decision have done.

⁶⁵ See *W v United Kingdom* (1987) 10 EHRR 29, para 73; *Bodén v Sweden*, n 42 above, para 28; *Pudas v Sweden*, n 42 above, para 30; *H v Belgium* (1987) 10 EHRR 339, para 40. The lateral move did not go unnoticed or uncriticised: see in particular the joint separate opinion of Judges Lagergren, Pinheiro Farinha, Pettiti, Macdonald, De Meyer and Valticos in *W v United Kingdom* and the opinions of Judge de Meyer in *Bodén v Sweden*, *Pudas v Sweden*, and *H v Belgium*. Usually the Court moved briskly over the obstacle but see *Salerno v Italy*, n 49 above, for an unusual deployment of the argument

⁶⁶ (1986) 8 EHRR 329.

⁶⁷ *ibid* para 192.

was cause for alleging non-compliance with the statutory compensation provisions.⁶⁸ The bigger question, concerning the right to the property in the first place, was simply left to one side by the majority of the Court. It was, however, the subject of a strong separate opinion by Judge Lagergren, which was joined by Judge Macdonald:

There is no doubt that the applicants' ownership of shares in the companies concerned constituted 'civil rights', which they could defend before the ordinary courts in the United Kingdom. These proprietary rights were extinguished, against compensation, through the operation of the [Aircraft and Shipbuilding Industries Act 1977]. The applicants were left with no remedy against the basic provisions contained in the Act; in particular, they had no possibility of obtaining a review by a national court of many of the arguments now presented to the Court at Strasbourg. It was merely the application of the 1977 Act that was left to be determined – in the last resort – by the Arbitration Tribunal and other national courts.

Thus, up to 17 March 1977, when the third Aircraft and Shipbuilding Industries Bill received the Royal Assent, the applicants could defend their proprietary rights against all kinds of interference. However, when the ultimate interference occurred, namely when the proposal to deprive them of their shares passed into law, the United Kingdom took away, in those respects that are relevant in this case, the jurisdiction of its courts with regard to the very existence of the applicants' proprietary rights and did not even confer on those courts any jurisdiction over the content of the statutory compensation provisions.

Judge Lagergren thought 'this abrupt and serious limitation of the right of access to courts' to be 'premature and unacceptable and ... not consistent with a fair interpretation of Article 6 of the Convention'.⁶⁹ But he joined the majority 'in view of the Court's recent judgment in the case of *James and Others* ...'.⁷⁰ The die was now firmly cast.

A succession of later cases dealing with the point followed the *James* rather than the *Ashingdane* approach. In *Powell and Rayner v United Kingdom*,⁷¹ for example, the applicants were complaining, *inter alia*, about their inability effectively to assert their civil right to sue for nuisance in respect of the noise problems caused to them by aircraft using Heathrow airport, which was situated close to their respective properties. Just as in *Ashingdane*, the problem lay not in the extinction of their right of action but in the drastic curtailment to it that had been effected by primary legislation; under section 76(1) of the Civil Aviation Act 1982, nuisance was only available as a remedy against an otherwise rule-abiding aircraft if it could be shown that its flight was not reasonable 'having regard to wind, weather and all the circumstances of the case.' The Commission considered that the case did not involve any civil right and it was for this reason that the applications were declared manifestly ill-founded under Article 6(1).⁷² This finding was rather curious and when the issue reached the Court (disguised as an Article 13 point which for present purposes need not detain us), the applicants relied on *Ashingdane* to

68 *Ibid* see further the Report of the Commission Series B, vol 85, paras 451–478 and the oral submission to the Court of the government's lawyer, Robert Alexander QC, 519.

69 n 66 above, 401 citing, *inter alia*, his own concurring Opinion in *Ashingdane v United Kingdom*, above n 43, though as suggested above it would have been possible, as with *James*, to have justified this denial of access to the courts as compatible with Art 6(1); see text above following n 58.

70 But see the short joint dissenting Opinion of Judges Pinheiro Farinha and Pettiti which appears to be rooted in the same point: n 66 above, 401–402. It is developed at greater length in Judge Pettiti's separate dissenting Opinion: n 66 above, 402–405.

71 (1990) 12 EHRR 355.

72 See *Powell and Rayner v United Kingdom* (1989) 12 EHRR 287, 293. For the admissibility rulings see *Powell v United Kingdom* (1987) 9 EHRR 241; *Rayner v United Kingdom* (1987) 9 EHRR 375; (1986) 47 *Decisions and Reports* 5. *Baggs v United Kingdom* (1987) 9 EHRR 235 arose from the same set of facts.

establish that they did indeed have a civil right which was rooted in the common law of nuisance, that this right had been curtailed by section 76(1), and that the statutory limitation that was there imposed on their common law action infringed the principles set out in *Ashingdane* 'in that it did not pursue a legitimate aim, it placed a disproportionate burden on the applicants and, as a result, it destroyed the very essence of their abovementioned common law right'.⁷³ This would have been an entirely logical way in which to have approached the Article 6(1) dimension to the case, but it was no longer as simple as this after *Lithgow* and *James*.

The Court answered the submission by asserting that,

the effect of section 76(1) is to exclude liability in nuisance with regard to the flight of aircraft in certain circumstances, with the result that the applicants cannot claim to have a substantive right under English law to obtain relief for exposure to aircraft noise in those circumstances. To this extent there is no 'civil right' recognised under domestic law to attract the application of Article 6(1).⁷⁴

Rather than confront directly the fact that what was in issue here was an Act of Parliament delimiting a common law right, the Court preferred to deploy the concept of 'recognition' as a route to finding that no 'civil right' had been established on the facts before it. Of course this was convenient because it meant that the Act of Parliament was not being directly challenged. But because it was absurd to suggest that the applicants' civil right to sue in nuisance had been removed, the Court found itself having to define the non-recognised civil right before it with extreme particularity as a right to sue in certain narrowly defined circumstances.

The effect of this ruling, as in both *James* and *Lithgow*, was to permit the content of a civil right under Article 6(1) to be effectively shaped by the domestic law of each and every member state. The concept of what was a 'civil right' may have remained autonomous and the exclusive responsibility of the Strasbourg organs, but whether such rights were 'determined' (unlocking Article 6(1) protection) or not 'determined' (thereby excluding Article 6(1)) had become after these cases fully a matter for the national authorities, subject to a distinction being drawn between limitations that were procedural and those that were substantive. In the Commission rulings on these cases, the substantive/procedural distinction hinted at in *James* was picked up and used⁷⁵ to justify the decision against the applicant taken at the admissibility stage: the applicant had no 'substantive right to compensation' since the purpose of the legislation was 'to exclude generally any possible compensation claims for trespass and nuisance and not just to limit jurisdiction of civil courts with regard to certain classes of civil action.' The latter kind of limitation, it is suggested, would have been susceptible of Strasbourg review. As we shall see this distinction was to become central to later Commission decisions on Article 6(1), not least in *Osman* itself.⁷⁶ As the Court put it in *Powell and Rayner*, '[a]ccess to the domestic courts is available to any person who considers that he has a cause of action in nuisance under English law. If a question of the application of section 76(1) arises, it will be for the courts to decide.'⁷⁷ This is precisely the kind of formalism that was so unequivocally rejected in *Ashingdane*. But the Court in *Powell and Rayner* was now unanimous, with the

⁷³ *Powell and Rayner v United Kingdom*, n 71 above, para 35.

⁷⁴ *ibid* para 36, citing *Lithgow v United Kingdom*, n 66 above.

⁷⁵ At least in *Rayner*, n 72 above, 47 Decisions and Reports 5, 15.

⁷⁶ See n 118 below.

⁷⁷ *Powell and Rayner v United Kingdom*, n 71 above, para 36.

flickers of dissent that had flared up in *James* and *Lithgow* having been fully doused.

Substantive and procedural rights

There the issue rested for a number of years before being vividly revived in *Fayed and Fayed v United Kingdom*.⁷⁸ Significantly from the perspective of *Osman*, the question in this case related not to a statutory inhibition of a civil right but rather to a restriction firmly rooted in the common law. The applicants complained that they were unable effectively to vindicate their civil right to their respective reputations in relation to statements made about them in a report into their financial affairs. This report, which was published, had been written by Department of Trade and Industry inspectors and it had followed a long investigation into the applicants' business affairs which had been initiated by the Secretary of State for Trade and Industry acting under statutory powers. The problem for the applicants lay in the fact that any defamation suit that they might wish to launch against the inspectors would be bound to fail because of the well-established common law rule that such reports attracted either absolute or qualified immunity from any such suit.⁷⁹ So was this a case of the wrongful denial of access to a court?

The European Court decided unanimously that it was not, but for reasons which are not easy to summarise. The Court was clearly aware that it was concerned in the case with the civil right to a reputation rather than with the right of access to a court as such.⁸⁰ Moreover, citing *James* and *Powell and Rayner*, it identified the right to a court as extending 'only to "contestations" (disputes) over (civil) "rights and obligations" which can be said, at least on arguable grounds, to be recognised under domestic law'. It also accurately quoted these cases as authority for the proposition that Article 6(1) 'does not in itself guarantee any particular content for (civil) "rights and obligations" in the substantive law of the Contracting States'.⁸¹ Viewed solely by reference to these authorities, it was clear that the applicants had no chance of successfully establishing a breach of Article 6(1); domestic law had been clearly so constructed that the applicants' right to their reputation had been effectively and unarguably 'derecognised', at least as far as the facts before the Court were concerned. But clear though they were, these dicta must have been unsettling for the Court in *Fayed*, seeming to involve giving rather too much power to national authorities. The circumstances moreover had less of the political potency of *James* and *Powell and Rayner*; as we have seen, these latter cases had directly involved legislative manifestations of parliamentary sovereignty whereas what was in issue here was judge-made law, albeit a law that had developed in a highly political context. Threshing around for a way of diluting the influence of the dicta to which it had just committed itself, the Court picked-up on a couple of threads in its earlier jurisprudence which it considered to be reflective of a more assertive Strasbourg position. It has to be said that the Court did this in a way which showed little appreciation of context or (more importantly) of the relationship between these formulae and other parts of the Court's jurisprudence, with results that were problematic to say the least. In particular in the second of these threads, it recovered and misapplied the *Ashingdane* 'right to a court' dictum

⁷⁸ (1994) 18 EHRR 393.

⁷⁹ cf *Agee v United Kingdom* (1977) 7 *Decisions and Reports* 164.

⁸⁰ *Fayed v United Kingdom*, n 78 above, paras 54 and 58.

⁸¹ *Ibid* para 65.

in a way which has turned out to be more damaging (from a conceptual point of view) than if the case had been entirely forgotten.

As far as the first of these threads is concerned, we have already noted how in *James* the idea of a European obligation not to impose Article 6(1)-based 'substantive' rights on national states had crept into the case-law.⁸² As we have seen, the Court had already used this phrase when denying an Article 6(1) right to the applicants in *Powell and Rayner*.⁸³ Now, spotting its possible utility, the respondent government in *Fayed* argued that the law on privilege amounted to what could credibly be described as a substantive limitation on the applicants' civil right and as it was such a thing, it was therefore a limiting action in which a member state could engage without attracting Article 6(1) stricture. The applicants inevitably responded that the common law before it was in fact merely a 'procedural' limitation and as such was not the kind of 'substantive' limitation that could drive a set of proceedings beyond the protection of Article 6(1), since merely 'procedural' barriers to access to a court continued to attract Article 6(1) scrutiny even if 'substantive' limitations on the self-same rights did not. It cannot be emphasised enough just how wrong-headed this distinction is. As we have seen, the idea of a 'substantive' right was never intended to be developed in opposition to the notion of a 'procedural' right. Insofar as it had any meaning in the Article 6(1) context, it was merely a reference to the fact that, in *James*, *Lithgow* and *Powell and Rayner* the Court had accepted that domestic jurisdictions could indeed derecognise 'civil rights', and therefore deprive those whose civil rights were lost in an abrupt way of access to the procedural safeguards of Article 6(1).

But, despite having applied *James* and *Powell and Rayner* in an orthodox way as described above, the Court also contrived to fall in a half-hearted way for the seductiveness of the substantive/procedural distinction: the national authorities could 'derecognise' substantive rights but could not willy-nilly impose 'procedural' limitations on them without falling foul of Article 6(1):

Whether a person has an actionable domestic claim may depend not only on the substantive content, properly speaking, of the relevant civil right as defined under national law but also on the existence of procedural bars preventing or limiting the possibilities of bringing procedural claims to court. In the latter kind of case Article 6(1) may have a degree of applicability.⁸⁴

It is not obvious here what is meant by a 'degree of applicability.' As the Court admits, '[i]t is not always an easy matter to trace the dividing line between procedural and substantive limitations of a given entitlement under domestic law. It may sometimes be no more than a question of legislative technique whether the limitation is expressed in terms of the right or its remedy.'⁸⁵

Having recognised this, why did the Court not explicitly jettison the distinction completely? Instead it went on to assert that '[i]n the present case [it is not] necessary to settle the question of the precise nature of the defence of privilege for the purposes of Article 6(1), since it is devoid of significance in the particular circumstances.'⁸⁶ Apparently this was because if 'the Court were to treat the facts underlying the complaints declared admissible by the Commission as raising a

⁸² See text at n 61.

⁸³ For the Commission's contribution, see n 72 above.

⁸⁴ *Fayed v United Kingdom*, n 78 above, para 65. See *ibid* para 63 of the Commission Opinion for the argument of the government on the point. And see also Mrs J. Liddy's concurring opinion which stresses the substantive/procedural distinction.

⁸⁵ *ibid* para 67 of the Judgment.

⁸⁶ *ibid*.

substantive, rather than a procedural, complaint going to the right to respect for private life under Article 8 of the Convention – as it has jurisdiction to do – the same central issues of legitimate aim and proportionality as under Article 6(1) would be posed.⁸⁷ So does the procedural/substantive distinction never apply? Or does it not apply only where the relevant substantive civil right can also be repackaged under some other Article of the Convention? If the allegedly 'substantive' right to a reputation were not also capable of falling within Article 8, would it then be critical for Article 6(1) purposes whether the national limitation on that right were or were not procedural rather than substantive? It would seem so, even though the Court has itself recognised the distinction to be largely formalistic. Perhaps this is after all what is meant by a 'degree of applicability'.⁸⁸

In other passages of its judgment in *Fayed*, the Court leaves to one side the complexities of the procedural/substantive distinction and seeks to meet the problem of national subjectivity inherent in the threshold arguability test it has earlier applied in a different, half-traditional kind of way. This is the second thread of authorities referred to above. The Court drew upon a Commission admissibility decision in order to assert – without seeking to make any connection with the passages quoted above – that while '[c]ertainly the Convention enforcement bodies may not create by way of interpretation of Article 6(1) a substantive civil right which has no legal basis in the State concerned,' at the same time

it would not be consistent with the rule of law in a democratic society or with the basic principle underlying Article 6(1) – namely that civil claims must be capable of being submitted to a judge for adjudication – if, for example, a State could, without restraint or control by the Convention enforcement bodies, remove from the jurisdiction of the courts a whole range of civil claims or confer immunities from civil liability on large groups or categories of persons.⁸⁹

These comments then led the Court to a formulation of general principles on when access to a court may be legitimately denied or restricted which the Court acknowledged was derived from *Ashingdane*.⁹⁰

As we have seen, this analysis had a coherent jurisprudential pedigree behind it, from *Golder* to *Ashingdane*. However, and we can now see that this was a crucial omission from the perspective of later cases, the Court did not make explicit that these 'access to a court' principles only applied as a fallback European position in those cases in which the primary threshold test could not be successfully negotiated by applicants, and therefore (for this reason alone) the ordinary Article 6(1) principles could not be made to apply. The impression given in *Fayed*, therefore, despite its earlier restatement of the arguability test, is that the entitlement to a court was a free-standing right in Article 6(1).⁹¹ This heretical position was given added weight by the general statement with which the Court concluded its summary of the case law on the subject, when it said that '[t]hese principles reflect the process, inherent in the Court's task under the Convention, of striking a fair balance between the demands of the general interest of the community and the requirements of the protection of the individual's fundamental rights.'⁹² This may

⁸⁷ *ibid*, citing *Handyside v United Kingdom* (1979) 1 EHRR 737 on the jurisdictional point.

⁸⁸ See the very clear concurring (but on this point highly critical) Opinion of Judge Martens in *Fayed v United Kingdom*, n 78 above, 437–9. And see the somewhat critical comments of Bratza, n 14 above, 9.

⁸⁹ n 78 above, para 65, citing *Dyer v United Kingdom* (1984) 39 Decisions and Reports 246.

⁹⁰ *ibid* also citing *Lithgow v United Kingdom*, n 66 above.

⁹¹ Compare the clarity of the Commission Opinion on the point, *ibid* paras 71–6, where *Ashingdane* (n 43 above) is helpfully relied upon.

⁹² *Fayed v United Kingdom*, n 78 above, para 65.

be appropriate language to use in relation to substantive Convention rights, but it is a little grand for what conceptually ought to be no more than a subsidiary test governing the application of Article 6 to other civil rights in certain circumstances. In *Fayed* itself this half-treatment of the right of access to a court as a substantive right did not cause problems because the applicants had indeed failed to surmount the threshold test for entry to Article 6(1): their attempt to vindicate their reputations was indeed unarguable in domestic law in light of the common law on privilege. It followed therefore, though the Court was never explicit about any of this, that on the authority of *Ashingdane* it was right to treat the case as one involving the right of access principles. (Hints of the paradoxes that later engulf this jurisprudence are however to be found in an early form in *Fayed*: we should note in particular the assertion that judicial review was available to the applicants if they had wanted to challenge the inspectors' damning indictment of their reputations, and that therefore the denial of their access to a court was less severe than they suggested;⁹³ but if judicial review was available, why did the applicants not have an arguable case that their civil right continued to be recognised in domestic law?)

If the two-fold threshold test in *Ashingdane* had been remembered and applied, the arid nonsense over procedural and substantive limitations could easily have been avoided, and the apparently simultaneous application of what were in fact either/or threshold/European fallback tests would also not have been undertaken. Under *Ashingdane* the first question would have been the simple one, whether the applicants had an arguable case that the interference with the exercise of their (civil) rights was unlawful, to which the answer would plainly have been no (the law on privilege was clear and no malice had been alleged). This conclusion would then have led to the second question, which would have been concerned with whether the degree of access to the courts afforded under the national law was sufficient to secure the individual's 'right to a court', having regard to the rule of law in a democratic society. The *Ashingdane* principles would have come into play in answering this second question in exactly the way they did in *Ashingdane*, but without these long detours into the worlds of substantive, procedural and Convention/non-Convention rights. The problem in the Court's case-law both before and in *Fayed* stems from the failure always clearly to see the right of access to a court as a fallback European safeguard against domestic abuse rather than as a substantive Article 6(1) protection in its own right.⁹⁴ As though this problem were not bad enough, we should also remember that the position continued to be further complicated by the replacement in *James* of the idea of unlawfulness in the *Ashingdane* threshold test with a kind of quasi-positive version of the same, the woolly notion of a 'recognised' or 'unrecognised' civil right of a 'substantive' (as opposed to a 'procedural'?) nature.

The right to a court

Of all the analytical errors in *Fayed*, perhaps the most crucial was the simultaneous application of both the arguability threshold and the right-of-access test as if the latter were not dependent on the former being answered in a negative

⁹³ n 78 above, para. 78.

⁹⁴ Except in so far as the right can be viewed as a residual substantive right applicable when access to the procedural safeguards set out in Art 6(1) has been denied by national law: see text at n 53 above.

way. If an applicant had an arguable case, the right thing to do was to move straight into the substance of Article 6(1). If he or she did not have such a case, the right approach was then to proceed to check the legitimacy of that national denial of a court by reference to the *Golder/Ashingdane* etc line of cases. What the Court could not coherently do and should not have done was to ask the access-to-a-court question only after the threshold question had been answered affirmatively. This was to get the law exactly the wrong way round, with the Court viewing the right of access as a right that becomes available only when the arguability question is answered positively, when in fact it should only be resorted to (as a fallback against over strict exclusionary rules at the national level) when that question is answered negatively. So after *Fayed*, case after case is decided where either the issue of right of access is dealt with after the arguability threshold has been negotiated,⁹⁵ or where the issue never arises because applicants are defeated on the threshold point from which they are permitted no escape route.⁹⁶ The Court seemed perfectly happy with this development, now often analysing Article 6(1) under the general rubric of 'right of access to a court,'⁹⁷ and generally loosening the bonds that connected the notion of access to a court with its previous, more clearly defined case law. In *Stubbings and Others v United Kingdom*,⁹⁸ the Court even goes so far as to dispense altogether with the arguability threshold in a way which has added a further layer of complexity to the case-law. The facts were deceptively simple, and this indeed may be what caused the problem. The applicants had desired to sue certain persons in respect of alleged childhood sexual abuse, but had been foiled by the application of the Limitation Act 1980, which, as interpreted by the House of Lords, had effectively blocked all their claims.⁹⁹ The case was manifestly similar to *Ashingdane*. The applicants' case was unarguable under domestic law and so everything should have hinged on whether the relevant national rules which produced this result were themselves excessive and disproportionate under the European fallback position which had been explicitly applied in *Ashingdane* and implicitly but sloppily deployed in *Fayed*.

Somewhat unexpectedly, the Court neither applied the threshold test nor considered the European fallback position as an alternative to its application. Instead, building on the vague generalities to be found in the post-*Fayed* case law, it contented itself with recalling 'that Article 6(1) embodied the "right to a court", of which the right of access, that is the right to institute proceedings before a court in civil matters, constitute[d] one aspect.'¹⁰⁰ But it did not then add, as it had done in *Fayed*,¹⁰¹ that this right to a court 'extends only to "contestations" (disputes) over (civil) "rights and obligations" which can be said, at least on arguable grounds, to be recognised under domestic law; [Article 6(1)] does not in itself guarantee any particular content for (civil) "rights and obligations" in the

95 As in *Société Leveau Prestations v France* (1996) 24 EHRR 351.

96 See *Hamer v France* (1996) 23 EHRR 1, where however the judgment is very particular to the circumstances that there arose. Much clearer are *Gustafsson v Sweden* (1996) 22 EHRR 409; *Maxson and van Zon v Netherlands* (1995) 22 EHRR 491; *Andersson v Sweden* (1997) 25 EHRR 722; *MS v Sweden* (1997) 28 EHRR 313. For an intellectually rewarding concurring opinion (by Judge de Meyer) see *Rolf Gustafsson v Sweden* (1997) 25 EHRR 623.

97 As in *Ortenberg v Austria* (1994) 19 EHRR 524. To similar effect is *Fisher v Austria* (1995) 20 EHRR 349.

98 (1996) 23 EHRR 213. An extract from the Commission Opinion is at *ibid* 225-231. cf *Tolstoy Miloslavsky v United Kingdom* (1995) 20 EHRR 442.

99 *Stubbings v Webb* [1993] AC 498.

100 *Stubbings v United Kingdom*, n 98 above, para 48.

101 *Fayed v United Kingdom*, n 78 above, para 65.

substantive law of the Contracting States.¹⁰² Instead the Court in *Stubbings* plunged straight into the question of whether the restriction on the access to the court in this case was legitimate under Article 6(1), finding eventually that it was. There was no recognition of the fact that a threshold test even needed to be applied, much less that the right of access issue only arose if that threshold test was failed by the applicant. Instead, the right of access to a court was all that the Court felt needed to be addressed.

With this inelegant finesse, the right to a court made its final jump from being the fallback assertion of European oversight of the threshold question identified in *Ashingdane* into the (implied) substance of the Article itself.¹⁰³ After half-hinting as much in *Fayed*, the guarantee of access to a court had completed its return from near oblivion after *James* and *Lithgow* and in so doing had taken on a new and vigorous life, not as the qualified guarantor against over-rigorous national rules restricting access to Article 6 which it had once been but, rather, as a separate (but implied) right guaranteed by Article 6(1) itself. It had shed its *Golder* cover and become a substantive thing-in-itself. With *Fayed* and *Stubbings*, the Court set out on a path which was inevitably to lead, sooner or later, to an *Osman*-style decision. For though the Court in *Stubbings* had managed wholly to ignore the threshold point, the requirement of an arguable right was too embedded in the jurisprudence of Article 6(1) to be expelled entirely as a result of having merely been omitted from a single case.

The collapse into incoherence

Even if it had noticed the threshold test, the Court in *Stubbings* would have been right to have analysed the right of access cases since the threshold would not have been passed on the facts before it; as in *Fayed* the applicants had no hope in domestic law. In neither *Fayed* nor *Stubbings* was it necessary to address the question of what would happen when the traditional threshold test of arguability could indeed be successfully negotiated by an applicant in a case which also superficially appeared to require a positive application of this new, apparently substantive, Article 6(1) 'right to a court'. We have stressed more than once in this Article how nonsensical this inquiry is. It is perhaps just about forgivable to say once again that if the threshold test is passed, then the substance of Article 6(1) as it appears in the Convention should be applied, ie the rights that Article 6(1) actually guarantees, namely, to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. If the test is not passed then the right of access case-law should be deployed. But the threshold question should not and could not be a *sine qua non* for application of the right of access to a court under Article 6(1). It is helpful to view the issue in the abstract before turning to the recent cases. How could the right of access to a court be coherently discussed in a case in which it had already been found that the applicant had an arguable right before the courts under domestic law? Surely the right to a court was bound to be present where a case had been found to be arguable? If this was the case, though, what was the point to a supposedly substantive right in

102 *ibid* citing *James v United Kingdom*, n 55 above and *Powell and Rayner v United Kingdom*, n 71 above.

103 A case similar in reasoning which relied on *Stubbings* is *The National & Provincial Building Society, the Leeds Permanent Building Society and the Yorkshire Building Society v United Kingdom* (1997) 25 EHRR 127. See also *Canea Catholic Church v Greece* (1997) 27 EHRR 521.

Article 6(1) which merely duplicated the Article's threshold question? If it was not the case, how could a Strasbourg application ever be both arguable in court for one purpose (the threshold test) and not properly arguable in court for another (application of the supposed Article 6(1) substantive right of access) at the same time and on the same facts? In *McGinley and Egan v United Kingdom*,¹⁰⁴ the Court contrived simultaneously to apply the substance of Article 6(1) with the implied 'right to a court' in a way which may have avoided this difficulty, but which also inevitably suggested that not even the judges now knew the difference between the express and implied parts of Article 6(1), much less how they related to the arguability threshold which in the case before them had been effortlessly negotiated.¹⁰⁵ Having conflated so many of its previous principles into an ill-judged assortment of contradictory rules to be simultaneously applied, conceptual disaster could not be long postponed. It came twice, in quick succession, in 1998.

Tinnelly and McElduff v United Kingdom

The 'civil right' in issue in the first of these cases, *Tinnelly and McElduff v United Kingdom*,¹⁰⁶ was held by the Court to be 'the right not to be discriminated against on grounds of political belief or political opinion in the job market including ... when bidding for a public works contract or sub-contract.'¹⁰⁷ This right was guaranteed by statute in that part of the United Kingdom, Northern Ireland, in which the fact situation for both applicants in *Tinnelly and McElduff* arose.¹⁰⁸ The interference with the civil right was allegedly the action of the Northern Ireland Electricity Service (in *Tinnelly*) and of a branch of the Northern Ireland government (in *McElduff*). In both cases, a framework of law existed whereby the applicants' complaints could be adjudicated upon. Under section 42(1) of the relevant legislation, however, the Act was stated not to 'apply to an act done for the purpose of safeguarding national security or of protecting public safety or public order.' Under subsection 2 of the same section, '[a] certificate signed by or on behalf of the Secretary of State and certifying that an act specified in the certificate was done for a purpose mentioned in subsection (1) shall be conclusive evidence that it was done for that purpose.'

Armed with these legislative weapons, the respondent government's lawyers did not let the threshold question go by default on this occasion. They contended that the effect of section 42 was that there was no substantive right in national law upon which the applicants could arguably rely. The European Court replied by denying the government's characterisation of the effect of section 42: 'Whether or not the act of refusing the contracts and the security clearance necessary for obtaining them was an act done for the purposes of protecting, *inter alia*, national security is a matter which can properly be submitted for examination by a court or tribunal.'¹⁰⁹ Furthermore the legality of the certificates issued under section 42(2) 'was open to challenge [before the courts] at the very least on grounds of bad faith.'¹¹⁰ For these reasons the Court concluded that Article 6(1) was applicable to

104 (1998) 27 EHRR 1.

105 *ibid* paras 85–86, 90; cf *Vasilescu v Romania* (1998) 28 EHRR 241.

106 (1999) 27 EHRR 249.

107 *ibid* para 61.

108 The Fair Employment (Northern Ireland) Act 1976. The Fair Employment (Northern Ireland) Act 1989 was also relevant to the case.

109 n 106 above, para 62.

110 *ibid*. For the Commission's reasoning, see 268–280.

the proceedings which had been instigated by the applicants; this was presumably because (though the Court nowhere makes it explicit) there was indeed an arguable case here in relation to a substantive right that could be put before a domestic court and that consequently the right was one that was arguably recognised in national law.

Given that the threshold to Article 6(1) had been successfully negotiated, one would have expected the Court then simply to have asked whether the determination of the civil right in issue had been in a 'fair and public hearing' by 'an independent and impartial tribunal' established by law. This is where whatever anxieties there might have been about the restrictions on the courts' powers with regard to the evidence before it could and should have been ventilated.¹¹¹ Instead, beguiled by *Stubbings'* erroneous belief that the right to a court was an independent Article 6(1) right, the Court then proceeded to embark on an analysis of whether the restrictions on access to the court that were before it were proportionate and otherwise justifiable.¹¹² The Court's answer to this question was a unanimous negative. In the course of coming to this conclusion, as part of its proportionality analysis, the Court had naturally to draw attention to how dramatically restricted access to the national court was by the action taken by the secretary of state under section 42, and indeed it weighed this point heavily against the respondent government.¹¹³ In doing this, however, the Court could not help giving the impression that it was taking the very opposite position to that which it had adopted when considering the arguability of the applicant's claim at the threshold point in its analysis.

Following the same logic – inevitable once the right of access was applied not as a fallback to but only following a successful application of the arguability threshold – the applicants had also first argued that the restriction on evidence was arguable so as successfully to negotiate the threshold and had then, once into Article 6(1) with its supposed substantive right to a court, switched tack to emphasise the inflexible constraints that after all had clearly hampered the judicial inquiry into their case.¹¹⁴ Not to be outdone, the government's lawyers had also switched positions just as effortlessly when, having lost on the threshold point, they contrived to discover just how engaged had been the judicial role that they had previously so underplayed at the threshold stage.¹¹⁵ On the Court's final view of the law, the domestic framework had seemed flexible enough to allow an arguable case to be made to challenge the application of section 42, but not flexible enough for that arguable case to matter very much when it came to weighing the various relevant factors for the purpose of assessing whether the applicants had been wrongfully denied access to a court.

The *Osman* case

The decision in *Tinnelly and MacElduff* was pretty confusing and more than only apparently contradictory. Quite how difficult it was became fully appreciated only after the *Osman* case.¹¹⁶ The Article 6(1) issue in this decision, which as we saw in

111 See *Kerojärvi v Finland*, n 21 above; *Mialthe v France* (No. 2), n 22 above.

112 See in particular n 106 above, para 72, where *Stubbings* is the only authority explicitly cited.

113 See *ibid* at paras 74–75, 77–78.

114 cf *ibid* paras 64 and 65 with paras 59, 60 and 62.

115 cf *ibid* paras 58 (on the threshold point) and 66 (on the right of access to the court).

116 *Osman v United Kingdom*, n 4 above. See *Bellet v France*, European Court of Human Rights, 4 December 1995, for an interesting foretaste of the issues in *Osman*.

our introduction has caused such a furore, was analytically very similar to the issue in *Tinnelly and MacElduff*. The applicants' civil right was the right to sue the police for damages for negligence. The alleged interference with this right was the striking out of their claim in the English Court of Appeal on the ground that it disclosed no reasonable cause of action. This decision was arrived at because the relevant domestic authorities were able to convince the appeal judges that 'no action could lie against the police in negligence in the investigation and suppression of crime on the grounds that public policy required an immunity from suit.'¹¹⁷ Was this a breach of Article 6(1)?

Tinnelly and MacElduff has prepared us for the reasoning that the Court found itself compelled by past error to adopt. The first issue addressed concerned what the Court termed the 'applicability of Article 6(1)'.¹¹⁸ As we have seen, this has always been a way of describing the threshold test, and of course it was logical to start the Article 6(1) analysis at this point. It is unfortunate that in *Osman* the Court neither formulated the threshold test in any explicit way nor cited any relevant decisions which could have pointed the reader in the right analytical direction. The Court seems instead in this section of its judgment to have proceeded on the assumption that the point was implicitly understood by all. While perhaps true with regard to close scholars of the Article 6(1) jurisprudence (not a huge legal community), the elliptical nature of the Court's reasoning on this issue has not helped lawyers unschooled in the complexities of Article 6(1) in their efforts to understand this part of the judgment. Certainly neither set of counsel seems to have appreciated the significance of the threshold point, since their respective arguments (designed to deal with the 'substantive' right of access to the court) were counter-productive to their cases from the perspective of the threshold test: the applicants were arguing the inflexibility of the common law when they should have been stressing at this point its flexibility and the government was stressing its flexibility (to avoid the substantive argument) when it should at this juncture have been stressing its inflexibility.¹¹⁹

Nevertheless, despite these deficiencies, we can be fairly confident that the Court came to the correct decision on the (unarticulated) threshold test. It was surely right to say that 'the rule [in *Hill*] does not automatically doom to failure such a civil action [as that of the applicants] from the outset but in principle allows a domestic court to make a considered assessment on the basis of the arguments before it as to whether a particular case is or is not suitable for the application of the rule.'¹²⁰ Domestic cases supportive of this position were placed before the Court and acknowledged by it.¹²¹ It was the case therefore 'that the applicants must be taken to have had a right, derived from the law of negligence, to seek an adjudication on the admissibility and merits of an arguable claim that they were in a relationship of proximity to the police, that the harm caused was foreseeable and that in the circumstances it was fair, just and reasonable not to apply the exclusionary rule outlined in the *Hill* case.'¹²² It followed from this that 'the

117 These are the words of the Court: n 4 above, para 63.

118 *Osman v United Kingdom*, n 4 above, paras 133–140. For the Commission's approach, see Report of the Commission adopted 1 July 1997. See further the Commission decisions in *Z v United Kingdom* (1999) 28 EHRR CD 65 and *TP and KM v United Kingdom*, application 28945/95, both criticized by M. Lunney, 'Osman in Action Article 6 and the Commission Reports in *Z v United Kingdom* and *TP and KM v United Kingdom*' (2000) 11 KCLJ 119.

119 *Osman v United Kingdom*, n 4 above, paras 134 and 138.

120 *ibid* para 138.

121 See *ibid* paras 93–97 and 138.

122 *ibid* para 139.

assertion of that right by the applicants [was] in itself sufficient to ensure the applicability of Article 6(1) of the Convention'.¹²³

So far, so good, up to a point. What should now have happened, and what would have happened had the case been decided some years earlier, would have been that the substantive rights under Article 6(1) would then have been made available to the applicants. The question would have been the relatively simple one of whether the Osmans had had a 'fair and public hearing within a reasonable time by an independent and impartial tribunal established by law.' Meaningful questions would have been asked about the extent to which the procedure to which the applicants had been subjected could be described as fair, and this would have produced an outcome that might or might not have been satisfactory for them. For present purposes, the important point is not what the result would have been; it is that the Court's analysis would have had a clear contact with the litigious reality of the Osmans' case in that it would have dealt with and appraised from a procedural perspective what had been a real situation, namely one in which it was not remotely being disputed that the applicants had been before a court and in which the only relevant matter would have been the fairness with which they had been treated by the various courts before which they had put their case.

This is, however, not what happened. Just as in *Tinnelly and McElduff*, the Court was seduced by its earlier misguided case law into treating the right of access to a court as the key substantive right on the facts before it. This forced it into the inevitable contradictions that we have already seen so painfully exposed in *Tinnelly*. The domestic procedure, previously so open, was now suddenly to be condemned as inflexible: 'the Court of Appeal [had] proceeded on the basis that the rule provided a watertight defence to the police and that it was impossible to prise open an immunity which the police enjoy from civil suit in respect of their acts and omissions in the investigation and suppression of crime.'¹²⁴ The Court was now, it seemed, 'not persuaded by the Government's argument that the rule as interpreted by the domestic court did not provide an automatic immunity to the police.'¹²⁵ So, recalling that 'Article 6(1) embodies the "right to a court", of which the right of access, that is, the right to institute proceedings before a court in civil matters, constitutes one aspect',¹²⁶ the Court unanimously found that the domestic procedure in *Osman* 'constituted a disproportionate restriction on the applicants' right of access to a court' and as such was in breach of Article 6(1).

The reception of *Osman* into United Kingdom law

We have already noted at the start of this article the extent to which the *Osman* decision has mystified many members of the senior judiciary in this country.¹²⁷ In *Palmer v Tees HA*,¹²⁸ three Court of Appeal judges agreed with Lord Browne-Wilkinson's view, in the *Barrett* case, that *Osman* was 'extremely difficult to understand'.¹²⁹ In *Aldrich v Norwich Union*,¹³⁰ the Court of Appeal described the

123 *ibid.*

124 *ibid* para 150 and generally paras 148–53.

125 *ibid* para 152.

126 *ibid* para 147.

127 See nn 12–14 above.

128 n 13 above.

129 n 12 above.

130 n 13 above.

scope of *Osman* as uncertain. Morritt LJ's observations in *Jarvis v Hampshire County Council* are to the same effect.¹³¹ In *Kent v Griffiths*, Lord Woolf MR (as he then was) thought that it 'would be wrong for the *Osman* decision to be taken as a signal that, even when the legal position is clear and an investigation of the facts would provide no assistance, the courts should be reluctant to dismiss cases which have no real prospect of success.'¹³² His Lordship's concerns centred on the potentially adverse impact the decision would have on 'the new culture in litigation' introduced by the Civil Procedure Rules.¹³³ Lord Woolf thought however that, even though not technically binding, the decision was probably 'influencing the House of Lords to adopt a more restrictive approach to the exclusion of liability to categories of cases without first ascertaining their precise facts.'¹³⁴ This would seem undoubtedly to be the case.

In *Barrett v Enfield LBC*,¹³⁵ for example, there is a marked difference in the approach taken by the House of Lords, as against that of the Court of Appeal. The latter tribunal had accepted that reasonable foreseeability and proximity had been established before them, but had ruled that it was not just and reasonable to 'extend the common law duty of care by analogy to a local authority when it is performing its duties to protect and promote the welfare of children in its care.'¹³⁶ Argument in the House of Lords was heard in July 1998, but the decision was handed down after *Osman*, on 17 June 1999. It now turned out that these cases could be pursued after all. There was more scope for successful actions against local authorities acting in these circumstances than had been appreciated. Their Lordships were not explicitly 'following' *Osman*, but the spectre of the case hung over their various opinions.¹³⁷ In *W v Essex County Council*,¹³⁸ their Lordships once again intervened, post-*Osman*, to overrule a strike-out that had been upheld by the Court of Appeal in the pre-*Osman* era. The Strasbourg case was mentioned in argument before the House, but the single substantive opinion in the case, delivered by Lord Slynn of Hadley, did not explicitly refer to it. Instead their Lordships took the view that the case before them could not be described as unarguable. In *Arthur J S Hall v Simons*,¹³⁹ Lord Hoffman concludes his 21-page opinion dealing with the removal of the barristers' immunity in court by saying that the issue of Article 6 'does not arise'.¹⁴⁰ In *Jarvis v Hampshire CC; Phelps v Hillingdon LBC*,¹⁴¹ Lord Slynn ends his long judgment in a case which marked another sharp expansion of the liability of local authorities by observing, like Lord Hoffman in *Hall*, that he had not forgotten *Osman* but that the 'issue in that case does not arise'.¹⁴² Only in *Holland v Lampen-Wolfe*,¹⁴³ in which the traditional immunity enjoyed by states survived attack, were their Lordships inclined to refer liberally to the Strasbourg

131 [2000] ELR 36.

132 n 13 above, 485.

133 *ibid* 484.

134 *ibid*.

135 n 12 above.

136 As summarised by Lord Slynn at n 12 above, 203.

137 See especially Lord Hutton at n 12 above, 229 and Lord Browne-Wilkinson at n 12 above, 198–200.

138 [2000] 2 WLR 601.

139 [2000] 3 WLR 543.

140 *ibid* 578. Other of their Lordships were less disinclined to ignore the Convention altogether: see Lord Hope at 581, Lord Hutton at 605 and especially Lord Millett at 624. But, at least explicitly, the decision in the case is not driven by European considerations.

141 [2000] 3 WLR 776.

142 *ibid* 802. See also Lord Clyde at 808–809.

143 [2000] 1 WLR 1573.

case law, an enthusiasm perhaps not unrelated to the fact that this time it happened to be on their side.¹⁴⁴

Whatever the underlying reason, it is indubitably the case that after *Osman*, English law has made it much easier for many persons to sue public authorities in negligence. To this extent, the 'new culture in litigation', to which Lord Woolf referred in *Kent v Griffiths*,¹⁴⁵ has had inculcated into it a large dose of the 'right to adversarial proceedings', which the European Court of Human Rights has frequently described as the basic principle in Article 6(1).¹⁴⁶ The interesting question is whether the senior UK courts would have opened up the law of negligence in the way that they have without the impetus provided by *Osman*. There is no doubt that until only a few years ago, such a broadening of the law of negligence would have given rise to a most anxious debate about the consequences from a public interest perspective of opening the floodgates to successive waves of litigants. Had the matter been thought more appropriate for consideration in Parliament rather than in the law courts, it can be expected too that this issue would also have been very prominent. But who can dare to place such intangible public concerns on the weighing scales against a tragically unfortunate litigant with the rhetoric of human rights firmly behind him or her?

Conclusion

It is perhaps worth summarising again exactly how the jurisprudence of the European Court of Human Rights on Article 6(1) has ended in the kind of reasoning that we have seen displayed in *Osman* and *Tinnelly and MacElduff*. The root of the problem lies in the contrast between the French- and English-language versions of what appears in the English version of the Convention as the word 'determination'. Under the influence of the French equivalent's emphasis on a claim, the Court early on defined the idea of a *determination* of a civil right as something that necessarily involved some kind of dispute. Once this position was adopted, it naturally became important to work out which kinds of disputes could attract Article 6(1) protection and which could not. The answer to this problem which was developed in *Ashingdane* still seems by far the most satisfactory: the Article applied where it could arguably be claimed that an interference with a civil right was unlawful under domestic law, but where the local law was skewed against an individual so as to make the interference with his or her civil right inevitably legitimate under that law, then the European Court reserved the right to intervene, despite there being no arguable case. In the course of this article, we have called the first of these positions the threshold question and the second the European fallback position. Only the second of these positions should involve reliance upon the right of access to a court jurisprudence of the Court. (Of course the idea of an 'unlawful interference' with a 'civil right' is itself a difficult one for English private lawyers to grasp, but though obscure in its origins and foreign in its application, it has at least the virtue – not to be scoffed at in the present context – that it can be applied with reasonable clarity.)

Everything would have been fine with the subsequent case-law if this principled position had been consistently followed. Unfortunately it was not.

144 See Lord Hope of Craighead at *ibid* 1578 citing *Waite v Germany* (1999) 6 BHRC 499 and *NCF and AG v Italy* (1995) 111 ILR 153. To similar effect are Lord Cooke at 1578, Lord Clyde at 1581 and Lord Millett at 1588–1589.

145 n 13 above.

146 See, for example, *Van Orshoven v Belgium* (1998) 26 EHRR 55, para 41.

Under the political pressure (using this phrase broadly) of the high profile UK-based cases of *James* and then (to a lesser extent) *Lithgow*, the Court dropped the European fallback position in its entirety and rephrased the threshold question as one rooted in recognition rather than unlawfulness. The result was an apparent contradiction in the case law. On the one hand, the Court continued to emphasise its power to define what was a civil right independently of the meaning accorded to such rights in domestic law: there is a long case law to this effect, asserting the autonomy of the European Court in this regard, which we have not discussed here.¹⁴⁷ On the other hand, the Court seemed at the same time to be allowing the domestic authorities to regulate access to Article 6(1): the local power to 'recognise' (or more to the point to 'derecognise') civil rights effectively pre-determined when the safeguards in Article 6(1) would apply, with the supposedly autonomous power of definition of civil rights being simply irrelevant to this preliminary question. In seeking to reconcile these opposite positions, the Court's later case law was tempted into experimentation with various meaningless distinctions, such as between procedural and substantive barriers to the court and between barriers which infringe substantive Convention rights and others which do not.

Such dicta have merely confused further an already confused position. Their effect has been mischievous rather than fundamental. The really serious analytical catastrophe came when the Court sought another way of resolving the tension described above by recalling to the front-line of its reasoning the European fallback position it had jettisoned in *James* but now presenting it in a new guise, not as an alternative to the threshold question but as a substantive Article 6(1) entitlement in its own right, capable of being relied upon (with other substantive rights) only when that first threshold question was answered positively. So the right to a court became a right of which applicants could avail themselves if they could first show that they had a right which was arguably recognised in their domestic law. Like a new and particularly virulent strain of some dangerous germ, the right to a court had soon pushed all the explicit rights in Article 6(1) aside and begun to pose as the entirety of the protection that Article 6(1) had to offer civil litigants. Hence the otherwise very mysterious recent proliferation of a line of cases which had previously been composed of a few scattered and exceptional decisions.¹⁴⁸ And hence also the frequency with which the Court must now effectively find simultaneously that a case is arguable before the domestic courts (the threshold question) and that it is not so arguable (the European fallback position masquerading as a substantive right). However it is interesting to note that recently the Court appears to have come to regard a defeat for an applicant to be unobjectionable from a right of access to a court perspective if it is rooted in the absence of proximity and/or foreseeability rather than in the 'fair, just and reasonable' criterion as was the case in *Osman*. It may therefore be that *Osman* will come to be seen as a very high high-water mark in relation to the expansion of Strasbourg's remit.¹⁴⁹

It is perhaps fair to say that the Strasbourg case law discussed in this article does not show the European Court at its finest. In its favour, we should not forget that the Court has throughout been driven by that most laudable of motives, the doing

147 See van Dijk and van Hoof, n 20 above, 398–406.

148 See, for example, *Aerts v Belgium* (1998) 29 EHRR 50; *Pérez De Rada Cavanilles v Spain* (1998) 29 EHRR 109. And see also from the Commission *Z v United Kingdom*, n 118 above.

149 See *TP and KM v United Kingdom*, n 118 above (a Commission decision); *Bromiley v United Kingdom*, application 33747/96, 23 November 1999; *Powell v United Kingdom*, application 45305/99, 4 May 2000.

of justice to individual applicants in their arguments with their national authorities. It is this that has driven the Court to expand the remit of Article 6(1) and to develop the concept of the right of access to a court as a means of critically assessing the content of the domestic law before it. To the extent that the Court is to be accused of analytical uncertainty, this is as much a consequence of an over-reliance as an under-reliance on earlier cases. The Article 6(1) jurisprudence we have discussed reveals a bench of judges half-addicted to precedent. Earlier cases are frequently cited, and whole sections of past decisions are often applied to some new set of facts that has come before the Court. The judges have shown themselves very reluctant to disregard earlier cases, explicitly to jettison this or that dictum or to marginalise lines of reasoning that have been manifestly developed in error. Where a paragraph or sentence is perceived not to be helpful to the resolution of a current difficulty, it is merely left to one side and omitted without remark. Generally speaking, however, the Court has on the whole sought to adopt a holistic approach to its past utterances, making them all relevant wherever and whenever it can, so that even remarks that have been exiled in one case (or even a series of judgments) cannot be guaranteed not to return, sometimes quite out-of-the-blue, and sometimes in quite novel contexts, unrecognisable to their begetters.

This is not precedent as the common lawyer understands the term. The judgments of the Court never engage with past case-law in the way familiar to a student of the common law. There are no attempts to steer a path through earlier cases, identifying strengths and weaknesses and setting out clear, principled guidance for the future. In the 1980s, there used quite frequently to be helpful separate opinions in which individual judges made powerful, often profound, contributions to the case law, but these have trickled to a halt in recent years. Another disconcerting recent tendency has been the Court's habit of setting out tests that need to be applied without making explicit their nature, their roots in the Convention or their origin in the case law. This has given to some recent decisions the feel of a private conversation among *cognoscenti*, where all are presumed to know why certain formulae or phrases are being deployed. The exclusionary atmosphere thereby generated has done little to dilute the hostility which perceived gaps in the reasoning involved might already (in certain cases) have generated.

What then is to be done? The first and most obvious point is that this article has been about conceptual clarity rather than the results in any of the cases we have discussed. It is still possible for the European Court to restructure its reasoning in order to liberate itself from the paradoxes and contradictions into which its past errors have led it. There is a new Court and an opportunity therefore for a fresh start. The United Kingdom also faces a fresh start, in the form of the Human Rights Act, now fully in force. It is now unlawful for a public authority to act incompatibly with a Convention right,¹⁵⁰ and the notion of a public authority here is explicitly said to include the courts.¹⁵¹ Included among the Convention rights in schedule one of the Act is of course Article 6(1). So, with the Act now in force, it would seem that it will be the primary responsibility of the courts to ensure that '[i]n the determination of his civil rights and obligations ... everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law.' Crucially, under section 2 of the Act, the case law of the European Court will be relevant but not decisive when it comes to interpreting the rights set out in the Act. As Sir Nicolas Bratza has put it,

¹⁵⁰ Human Rights Act 1998, s 6(1).

¹⁵¹ *Ibid* s 6(3)(a).

what is ... clear is that, with the coming into effect of the Act, our courts will be obliged to examine with considerable care all limitations on the right of a litigant to have a determination of the merits of a claim, whether such limitations are statutory or judge-made and irrespective of whether the limitation is properly characterized in domestic law as substantive or procedural.¹⁵²

A brave approach would be to start entirely afresh. After all, the only reason that the idea of a determination came to include the need for a dispute in the Strasbourg case law was because of the Court's desire to reconcile the French- and English-language versions of the Convention. British judges have no domestic imperative to engage in such interpretive diplomacy; there is only one language accorded definitional priority in the Human Rights Act and it is English.¹⁵³ On this analysis, all 'determinations' of civil rights, in the sense of their extinction or radical truncation, would without any dispute dimension to them, fall to be protected by the safeguards set-out in Article 6(1). Rather than have an 'arguability' threshold question in Article 6(1), the domestic courts could apply the ordinary 'victim' test to weed out deserving from undeserving supplicants.¹⁵⁴ The impugned 'determination' would not necessarily need to be by a judicial body, but there would in such circumstances need to be access to such a body if Article 6(1) were to be satisfied; how much access could be determined as a separate issue.¹⁵⁵ Where the relevant determination was by Parliament, however, then section 4 of the Act would insulate the interference from Convention oversight, thereby removing the *James/Lithgow* problem (with all its complicating potential) from the ambit of the Act. On the other hand, if the determination were made up of a common law rule, then the matter would fall to be discussed as relating to the fairness of the applicant's hearing rather than the extent of his or her 'access to a court'.

The apparent simplicity of such an approach makes it particularly appealing. It would not, however, remove the entirely separate Article 6(1) questions relating first, to what is a civil right and, secondly, to the depth of judicial engagement that is required to make the determination of a civil right compatible with Article 6(1). These threads to the jurisprudence of the Strasbourg organs, not discussed in this article, would still need separately to be addressed. And the case law it does remove – on the threshold question – may be simply replaced by new complexities, relating to when a right can be said to have been sufficiently determined to attract Article 6(1) protection. For the word 'determination' is not unproblematic in English law, even when separated from its French Convention equivalent. Would the postponement of a civil right be enough (as in *Golder*)? What if the civil right as such is not extinguished but that aspects of it are? How serious must a partial erosion of a right be to warrant its being described as a 'determination'? Are all sub-legislative 'determinations' of civil rights to be given Article 6(1) protection as a matter of course even where the impugned interference is, apart from Article 6(1), perfectly lawful? What are the implications of this for, say, the making of discretionary delegated legislation, to which it may be that section 4(1) protection might well not be extended?

Of course it might be said that these complexities and (no doubt) confusions would at least have the advantage of having developed from a recognisably coherent starting point and that therefore they could be made the subject of

152 'The Implications of the Human Rights Act 1998 for Commercial Practice', n 14 above, 10.

153 The Act gives no guidance as to the status for interpretive purposes of the French language version of the Convention.

154 Human Rights Act 1998, s 7(1).

155 For the European approach, see *Fisher v Austria*, n 97 above. See, further, *Kingsley*, n 36 above.

ongoing correction, something at which the common law excels. On reflection, however, it would seem that in addition to the general victim requirement Article 6(1) does need its own threshold question, if its protection for all determinations of civil rights is not to prove too dogmatic and insistent. Even if this were not the case, there is the risk that domestic case law on Article 6(1) which strays too far from the Strasbourg paradigm will be vulnerable to challenge before the European Court, to which applications from suitably qualified victims may still be brought post-implementation. That Court has to apply the dual-language Convention even if British judges do not, and it is this dual-language version as interpreted in Strasbourg to which the British government remains committed as a matter of international law.

There is a case therefore for an approach to Article 6(1) in the British courts which adopts the best features of the Strasbourg case law and which then applies these in a principled way to meet particular local conditions. By way of conclusion it is tentatively suggested that such an approach would be achieved by restoring to primacy the set of principles first clearly set out in *Ashingdane*, and thereby offering Article 6(1) protection to all determinations of civil rights which raise an arguable case of unlawfulness under domestic law, with the 'right of access to a court' cases being brought into play only when this primary threshold test cannot be negotiated on account of some structural hostility in domestic law. The precise interconnection between the Human Rights Act and the application of Article 6(1) in UK law is beyond the scope of this Article. If the domestic law precluding the arguability of a victim's case is either part of or mandated by statute, the English court's analysis of the fallback European test would – if the bench were inclined to find an Article 6 violation – raise a section 4 issue concerning a declaration of incompatibility. On the other hand, if the law which made the victim's case unarguable in domestic law were part of the common law, then the domestic court would clearly be able, if the Article 6(1) analysis pointed in this direction, to develop the law so as to make it Convention-compatible. Such an approach is not entirely lost to the Strasbourg jurisprudence; flickers of an earlier commitment to rationality can be spotted from time to time.¹⁵⁶ The results of the cases discussed in the course of this article might well turn out to be the same, though the application of the 'fairness' test in Article 6(1) is different in several particulars from the 'access to a court' jurisprudence that has recently developed.¹⁵⁷ Care would no doubt need to be taken lest the fallback substantive right to a court be more sympathetic to applicants than the basic Article 6(1) safeguards. These are issues for another day. It is enough for the present to suggest that it is surely not too late for UK – and possibly also for European – law to be oriented along this direction.

156 See *McMichael v United Kingdom* (1995) 20 EHRR 205; *De Haes and Gijssels v Belgium* (1997) 25 EHRR 1; *Niderost-Huber v Switzerland* (1997) 25 EHRR 709; *Werner v Austria*; *Szoldos v Austria* (1997) 26 EHRR 310.

157 See van Dijk and van Hoof, n 20 above, 428–438.

Risk and Remoteness of Damage in Negligence

Marc Stauch*

The remoteness enquiry in negligence, which serves to exclude the liability of defendants for harmful consequences that their careless conduct caused, but for which it seems unfair to penalise them, has long been beset by uncertainty. Indeed, a common view is that this area of the law can be explained only by reference to diffuse considerations of 'legal policy'. This paper, however, argues that the remoteness enquiry represents a principled response to a problem that can arise, at a deep level, in ascribing a harmful outcome to the negligent exercise of individual agency. The relevant problem concerns the possible mismatch between the hypothetical 'risk-claim' in virtue of which conduct was faulty and the causal set that subsequently materialised for harm. The 'revised risk theory' that emerges from this analysis accounts for the majority of remoteness determinations. However, a few exceptions are also considered where 'policy', in a restricted sense, operates to extend or curtail a negligent agent's legal responsibility.

An enduring problem in the law of negligence is that of remoteness of damage, or, as it is sometimes termed, 'legal causation'.¹ This issue arises once the factual causation question of whether the defendant's breach of duty played a necessary part in the claimant's injury has been answered in the affirmative. For even then we may in some cases doubt whether, as a matter of justice, the defendant should be held liable. Typically, such doubts arise where the injury was one that would not have occurred without the faulty intervention of some other human agent, or was otherwise unpredictable in nature.

However, granted the need for some mechanism for limiting liability at this point, what form should it take? Two main, and conflicting, approaches have developed in the literature in answer to this question. First, there are those commentators who would reduce remoteness of damage entirely to an issue of 'legal policy'. In their view, the court will decide as a matter of law alone whether a defendant, whose fault played a necessary part in bringing about the claimant's injury, should pay. The result is not affected by any further consideration of the nature of the causal mechanism that led to such injury.²

In the second place, though, there are those who, taking their lead from Hart and Honoré's influential treatise on causation,³ would, while admitting that at least part of the remoteness issue is dictated by policy considerations, argue that much within

* Law School, The Nottingham Trent University. The author is grateful to Jonathan Griffiths, Peter Kunzlik, Simon Salzedo and Kay Wheat, as well as Hugh Collins, for their helpful comments on earlier drafts.

1 Some commentators would regard these terms as encompassing different issues – 'legal causation' going specifically to problems raised by voluntary third party interventions. However, the argument presented in this paper will suggest that the remoteness issue comprises a single enquiry.

2 These theorists, 'causal minimalists', can be subdivided according to whether they believe that judges are guided here by reference to some extra-legal goal (such as promoting economic efficiency, see eg G Calabresi 'Concerning Cause and the Law of Tort' (1975–1976) 43 U Ch LR 69), or simply decide each case on an *ad hoc* basis in line with their intuition as to what is fair and appropriate, see eg L Green 'The Causal Relation Issue in Negligence Law' (1961) 60 Mich LR 543.

3 H.L.A. Hart and A.M. Honoré, *Causation in the Law* (Oxford: OUP, 1959) – subsequent page references are to the 2nd ed, 1985. For an overview of the main features of Hart and Honoré's argument, see J. Stapleton, 'Law, Causation and Common Sense' (1988) 8 OJLS 111.

it also has a quasi-factual rationale. In particular, Hart and Honoré drew attention in their work to common usage of terms of causal attribution and argued that, in many cases where courts hold damage too remote, we would also refuse, as a matter of ordinary speech, to say that such conduct had 'caused' the damage in question. This will be true especially where, in addition to the fact of the defendant's prior faulty conduct, the harm would not have occurred without the voluntary act of a third party or some independent abnormal occurrence.

When Hart and Honoré's work first appeared, it tied in well with the 'direct consequences' rule used by the courts for deciding whether some consequence was too remote.⁴ However, soon afterwards came the decision of the Privy Council in *The Wagon Mound*,⁵ since when the dominant trend has been for courts to analyse problems of remoteness in terms of the foreseeable risk of harm. At first sight, such an approach might be thought to provide a third way between the positions of Hart and Honoré and the causal minimalists. The courts are able to get by without obscure concepts such as 'directness of damage' and yet are apparently restrained by principle (in the guise of 'foreseeability of risk') in attributing liability.

Nevertheless, given the scope in the cases for disagreement as to what risks were reasonably foreseeable, an increasingly common view is that the language of foresight is used not in an explanatory, but rather in a conclusory fashion. In short, the courts first decide to impose or deny liability, as the case may be, and then invoke foreseeability, which serves to mask the real basis of their decision. If this is true, the use of the reasonable foreseeability test can be regarded as a victory for the causally minimalist position with its stress upon the open-ended nature of judicial determinations as to remoteness.⁶

The aim of this article, however, is to argue that such a conclusion would be premature and that Hart and Honoré's fundamental intuition against causal minimalism was justified. Paradoxically, though, it is a suitably modified version of the risk theory (underlying the reasonable foreseeability test), which best supports this view. Not only does this 'revised risk theory' serve to explain the majority of remoteness cases, but it also provides a moral rationale (in a way that reliance on linguistic usage alone does not) of courts' determinations on this question.

Accordingly, the first section of this article looks in more detail at the nature, and defects, of both the 'direct consequences' approach, and its more modern counterpart, the reasonable foreseeability test. Recently, there has been an attempt by the courts to buttress the use of foreseeability by appealing to the scope and purpose of the legal duty whose infraction, by the defendant, has led to the plaintiff's injury. However, it will be argued that, though useful in a few cases, generally the latter approach simply collapses back into the use of reasonable foreseeability.

The second section of the article proposes a more concrete notion of risk than that so far expounded by the courts, and one that seeks to imbue the concept with some objectivity. Hitherto the courts, in asking whether a risk of harm is reasonably foreseeable, have tended to concentrate on the nature of the 'outcome harm' (an inherently manipulable concept) rather than on the further conditions required (together with the faulty conduct) for such harm to have arisen. In arguing for attention to be refocused upon the latter, the theory developed in this article

4 As applied in *Re Polemis & Furness, Withy & Co Ltd* [1921] 3 KB 560.

5 *Overseas Tankship v Morts Dock Engineering Co (The Wagon Mound)* [1961] AC 388.

6 In *'The Wagon Mound No 2: Foreseeability Revised'* (1967) Utah LR 197, 205-206, Leon Green referred to foreseeability as a 'delightful and useful fiction with no restrictions in itself ... [which] serves in every case to call forth a fresh judgment'.

reveals a strong factual component. In particular, the remoteness enquiry is linked to the issue of factual causation and shown to require a comparison of the causal set for harm, as it actually materialised, with the 'risk-claim' that (hypothetically) would have been made at the prior moment of breach of duty.

The final section of the Article considers certain categories of cases in which the revised risk theory itself gives way to extraneous 'policy' considerations. This encompasses both cases in which the liability of the defendant will be extended beyond the reasonably foreseeable risks of his conduct and, conversely, where it is restricted so that he escapes liability for some consequences that *were* reasonably foreseeable.

The development of the law on remoteness

The causation and remoteness enquiries in negligence

As a tort, negligence is not actionable *per se*. For a suit to succeed, it is not enough that the defendant was in breach of duty (in that his conduct posed an unreasonable risk to a legally recognised interest of the claimant). The faulty conduct must have actually resulted in harm to the latter.

In order to determine whether the conduct caused the claimant's loss in a factual sense, we ask if it played a necessary part in such loss. Usually this enquiry involves the application of the 'but for' test: would the claimant have suffered the loss in the absence of the defendant's faulty conduct? If the answer is no, then the conduct made the difference; if yes, it did not – the same harm would have befallen the claimant anyway.⁷ For example, in the case of *Robinson v Post Office*⁸ the claimant suffered brain damage as the result of an allergic reaction to a tetanus injection. The doctor who gave the injection failed (in breach of duty) first to give a test dose in order to check for such allergy. Nevertheless, given that the injurious reaction only manifested itself three days after the full dose, the court was satisfied that a test dose would not have revealed the claimant's susceptibility in the half-hour normally allowed between the two doses. Accordingly, it held that the doctor's breach of duty did not cause the brain damage.

The factual causation enquiry serves to eliminate faulty conduct by the defendant that turns out, on closer inspection, not to have played a necessary part in the harm. On the other hand, what it does not do is provide a satisfactory way of eliminating faulty conduct that *has* played such a part, but where, given the very unexpected turn of events subsequently (often including the faulty interventions of other human agents), it strikes us as unfair to impose liability on the defendant. This function is served by the enquiry into remoteness.

Until 1961 the prevailing view in England was that, to be recoverable (as opposed to too remote), the harm had to be a 'direct consequence' of the defendant's faulty conduct. Though never closely defined, this term roughly encompassed consequences that supervened upon the conduct without the subsequent intervention of an abnormal occurrence or the voluntary act of another human agent.⁹ A contingency of the latter sort would qualify as a *novus actus*

⁷ The 'but for' test fails to give a result in certain 'causal over-determination cases where, instead, the fuller NESS (Necessary Element in a Sufficient Set) test will need to be applied, see R. Wright, 'Causation in Tort Law' (1985) 73 Cal LR 1735.

⁸ [1974] 1 WLR 1176.

⁹ See Hart and Honore, n 3 above, 136 *et seq.*

interveniens, which, metaphorically, 'broke the chain of causation'. Importantly, however, once a harmful consequence was found to be 'direct', this also provided a *sufficient* basis for a finding of liability. The fact that the harm in question was not reasonably foreseeable – ie it involved the materialisation of a risk that a reasonable person would not have guarded against – was nothing to the point.

The working of the 'directness' rule is conveniently illustrated by the facts of the case that became virtually synonymous with it, namely *Re Polemis*.¹⁰ The SS *Thrasylvoulos* was unloading at Casablanca when stevedores, hired by the defendant, carelessly dislodged a plank which fell into her hold. The stevedores' action was faulty because of the foreseeable risk of causing impact damage to the vessel or her cargo. In fact an explosion occurred, caused by a spark igniting petrol vapour in the hold, and the ship was totally destroyed. The Court of Appeal, upholding the arbitrator, accepted that the risk of *this* damage occurring had not been reasonably foreseeable, but nonetheless found the defendant liable.

The unfairness of the *Polemis* rule

In neglecting the need for reasonable foreseeability of the harm in suit, the judicial approach in the *Polemis* case is intuitively a harsh one.¹¹ This is because of the feeling we have that the faultiness of the defendant's conduct was not really relevant to the damage.¹² Rather he is saddled with liability because we have fortuitously identified some other source of fault in his behaviour (the taking of a different and unreasonable risk that in the event did not materialise).¹³ Of course, it is sometimes said that, given the choice between an innocent claimant who has suffered injury and a defendant who has, at any rate, behaved faultily in *some* respect (as well as causing the injury in a factual sense), it is better that the latter should pay. Nevertheless, as Keeton has tellingly argued:

... [I]f it is relevant to take into account defendant's fault with respect to a risk different from any that would include the harm plaintiff has suffered, then would it not also be relevant to take into account his other faults as well? And would it not seem equally relevant to consider plaintiff's shortcomings? Shall we fix legal responsibility by deciding who is the better and who the worse person?¹⁴

More generally, it is important to keep in mind that, in contrast to an agent who deliberately brings a harmful state of affairs into being, negligent conduct does not go to the defendant's culpability as an *individual*. Instead, in determining the breach of duty question, the court asks whether, in the same circumstances, a reasonable person would have run the same risk of harm as the defendant. The conduct of the defendant is then judged against this standard to decide if he has behaved faultily by falling below it (ie by failing to advert to the risk or by responding inadequately to it). Accordingly there is no implication that the defendant himself was aware of the danger, or, if he was, able to avoid it. Rather, as Brudner has argued, he is penalised by virtue of having fallen short of a general ideal of rational agency:

10 *Re Polemis & Furness, Withy & Co Ltd* [1921] 3 KB 560

11 In fact, as suggested later (see n 67 below), the case was probably correctly decided on its facts.

12 Admittedly the conduct was, in factual terms, causally relevant as it comprised a necessary condition for the injury, but its faulty quality was not (as that derived from some other risk, not the one that materialised).

13 See also *Thurogood v van den Berghs & Jurgens* [1951] 2 KB 537 and A. L. Goodhart's commentary on it, 'The Imaginary Necktie and the Rule in *Re Polemis*' (1952) 68 LQR 514.

14 R.E. Keeton, *Legal Cause in the Law of Torts* (Columbus: Ohio State University Press, 1963) 21.

... [I]f I did not foresee the consequence but caused it in circumstances where the reasonably circumspect person would have foreseen it as a likely consequence of my act, ... then the consequence is mine in a [weak] sense. It is mine because my free agency is a capacity for rising above the here and now, for surveying the network of interconnected forces within which my action occurs, and for anticipating the action's probable effects. Because I as an agent (with the normal powers of *homo sapiens*) cannot disown this capacity, I must acknowledge partial authorship not only of the effects that I as *this* individual foresaw but also those that the reasonably circumspect agent would have foreseen as a serious possibility.¹⁵

It is evident that cases where the particular risk that materialised was not reasonably foreseeable are exactly those where the contemplation of that risk would not have caused a reasonable person to behave any differently from the defendant. Here even the partial authorship of the defendant for the outcome is absent.

The rise of the 'risk theory'

In the light of dissatisfaction with the workings of the 'direct consequences' rule, the rival 'risk theory' approach to remoteness developed.¹⁶ The theory's central tenet is that a defendant should be liable only for harm that is 'within the risk' that made his conduct a breach of duty. In short, we should ask if the risk of the harm in suit would have deterred a reasonable person from engaging in such conduct. Alternatively, and to put it into the language of reasonable foresight, was the risk of the harm 'reasonably foreseeable'? (To describe a risk as reasonably foreseeable is merely another way of saying that it was a risk that a reasonable person would not have run, and whose running by the defendant was faulty.¹⁷)

According to its adherents, the risk theory/reasonable foreseeability test¹⁸ enjoyed two main advantages over the 'direct consequences' rule. The first of these was that it eliminated the potential unfairness, discussed above, of imposing liability in respect of a risk that a reasonable person would have run. In the second place, though, the use of the risk theory was supposed to simplify the law on remoteness of damage. Instead of considering the defendant's negligence in the abstract (because of the general risk of harm attending his conduct) and then asking if the harm in suit was a sufficiently 'direct' consequence, the courts would be able to engage in a single unified enquiry. The same basis for finding the defendant negligent (*viz.* that his conduct created an unreasonable risk of a given harm) would also serve to define that harm for which he was to be liable.

Essentially it was this theory that the Privy Council embraced when, in 1961, it came to review the law on remoteness in *The Wagon Mound*.¹⁹ The facts of the latter case concerned a fire that badly damaged the claimant's wharf and a ship, the

15 A. Brudner, 'Owning Outcomes' (1998) 11 *Canadian Journal of Law & Jurisprudence* 89, 104. Brudner cites Hegel's *The Philosophy of Right* as the source of inspiration for this passage.

16 The theory is generally regarded as deriving from the majority judgment of Cardozo J in the New York case of *Palsgraf v Long Island RR*, 248 NY 339, 162 NE 99 (1928), and swiftly gained a number of influential academic supporters on both sides of the Atlantic, including Keeton (see n 14 above), A.L. Goodhart ('The Unforeseeable Consequences of a Negligent Act' (1930) 39 *Yale LJ* 449, and see n 13 above) and Glanville Williams ('The Risk Principle' (1961) 77 *LQR* 179).

17 Frequently the concept of reasonable foreseeability is confused with that of 'theoretical' foreseeability – i.e. the awareness of the bare existence of a risk. The latter (a component of the 'duty of care'), though necessary to ground reasonable foresight is not enough if, in all the circumstances, the risk was one that a reasonable person would have run; see further, Hart and Honore, n 3 above, 262–263.

18 Strictly speaking, the 'risk theory' is wider than the 'reasonable foreseeability test' in that it is applicable in relation to breach of statutory duty as well. However, the two may be treated as equivalent in the context of negligence.

19 *Overseas Tankships v Morts Dock Engineering Co (The Wagon Mound)* [1961] AC 388.

Corrimal, which was being repaired there. The fire started because the claimant, in the course of carrying out welding works on the *Corrimal*, ignited bunkering oil on the water, which had been discharged by the defendant's vessel. Undoubtedly the defendant had behaved faultily in discharging the oil. However, this was not because of the risk of the oil catching fire (the oil had a very high flash point and, according to the evidence, was thought impossible to ignite on water²⁰), but because of the risk of fouling the wharf. The Privy Council, in disapproving *Re Polemis*, held that the defendant was not liable in negligence. It was not enough that the fouling of the wharf had been a reasonably foreseeable risk of his conduct; the risk of fire too must have been so foreseeable for liability to follow. In the words of Viscount Simonds:

[Their Lordships] have been concerned primarily to displace the proposition that unforeseeability is irrelevant if damage is 'direct'. In doing so they have inevitably insisted that the essential factor in determining liability is whether the damage is of such a kind as the reasonable man should have foreseen. This accords with the general view thus stated by Lord Atkin in *Donoghue v Stevenson*: 'The liability for negligence, whether you style it as such or treat it as in other systems as a species of "culpa", is no doubt based on a general public sentiment of moral wrongdoing for which the offender must pay.' It is a departure from this sovereign principle if liability is made to depend solely on the damage being the 'direct' or 'natural' consequence of the precedent act. Who knows or can be assumed to know all the processes of nature?²¹

Problems with the theory post-*Wagon Mound*

Unfortunately, in the years since *The Wagon Mound*, the hope that the use of the 'reasonable foreseeability test' would make for fairer and simpler law in relation to the remoteness issue has proven illusory. One problem has been the refusal of many judges to accept the 'extensive' aspect of the test, that is the fact that, as well as limiting liability relative to the *Polemis* rule (in cases where the harm was 'direct' but not reasonably foreseeable), the risk theory may on other occasions extend liability in cases where the harm was reasonably foreseeable but 'indirect'.²² This refusal has robbed the reasonable foreseeability test of much of the simplifying potential it might have had. As this, though, is a problem that applies in equal measure to the revised risk theory to be outlined in the next section, it will be deferred for discussion.

However, the principal difficulty has been to find some objective method of defining the harm, the risk of which made the defendant's conduct negligent. In *Wagon Mound* itself, Viscount Simonds suggested that the risk of the 'kind' of harm in suit must have been reasonably foreseeable, and evidently regarded fire damage as different in kind from damage by fouling. Nevertheless, given that any event in the world can be described in many different ways, the question as to whether the harm was reasonably foreseeable will ultimately turn upon the specificity of the description applied by the judge. As Fridman and Williams comment, 'If the level of abstraction is set low enough, then almost nothing will be foreseeable. Vague generalisations will cover almost everything, whereas "burns" will not include "cuts"'.²³

20 This finding was made on the basis of restricted evidence led by the claimant, who was embarrassed by the prospect that, otherwise, his own act in continuing with the welding works after he spotted the oil, would be held to amount to contributory negligence – a complete defence in New South Wales at the time. (Arguably this fear was unjustified; see n 55 below.)

21 See n 19 above, at 426.

22 See the discussion of 'Policy Exceptions' below.

23 'The Atomic Theory of Negligence' (1971) 45 ALJ 117.

A number of subsequent cases bear out this point. Perhaps best known is the House of Lords' judgment in *Hughes v Lord Advocate*,²⁴ which concerned the liability of the Post Office for injuries sustained by a boy when a paraffin lamp, left by its workmen near to an open manhole, exploded on being knocked into the manhole. In unanimously finding for the claimant, even though the explosion had not been reasonably foreseeable, the House of Lords chose to focus on the general risk of burning injuries posed to children by the unattended lamp. As Lord Pearce commented:

The obvious risks were burning and conflagration and a fall. All these in fact occurred, but unexpectedly the mishandled lamp instead of causing an ordinary conflagration produced a violent explosion. Did the explosion create an accident and damage of a different type from the misadventure and damage that could be foreseen? In my judgment it did not. The accident was but a variant of the foreseeable.²⁵

On other occasions, however, the courts appear to have demanded that the precise details of the risk that materialised were reasonably foreseeable. For example, in *Doughty v Turner Manufacturing*,²⁶ a scalding-injury that occurred following a chemical eruption of boiling liquid was distinguished from the same injury if it had resulted from a splash. Only the latter was said to have been reasonably foreseeable. In similar vein is the case of *Tremain v Pike*, in which the claimant contracted Weil's disease from working on the defendant's rat-infested farm. Payne J held that though injury by rat-bites was reasonably foreseeable, Weil's disease – a barely known condition at the time – was not.²⁷

The 'scope of the duty' approach

More recently, the courts have deployed a different version of the risk theory, which at first sight has lent greater authority to their decisions on remoteness. In particular, they have sought to define that harm for which defendant is to be liable by appealing to the scope and purpose of the legal rule, which imposed the duty upon the defendant to behave carefully.²⁸

A good example of this 'scope of the duty' approach is provided by *Banque Bruxelles SA v Eagle Star*.²⁹ Here the claimant mortgagee advanced monies on a property that had been negligently overvalued by the defendant surveyors. When the mortgagor became insolvent, and the claimant sold the property, it recovered only a fraction of what it had lent, due largely to the collapse of the property market in the meantime. In its action against the defendant, it attempted to recoup the entirety of its loss on the basis that, had it known of the true value of the property, it would not have advanced any monies at all to the mortgagor. In finding that the defendant was

24 [1963] AC 837.

25 *ibid* 858.

26 [1964] 1 QB 518. This was a decision of the Court of Appeal, which also included Lord Pearce.

27 [1969] 3 All ER 1303. As Michael Jones has pointed out (*Textbook on Torts*, London: Blackstones Press, 6th ed, 1998, 242), 'If the question had been: Was illness from some rat-transmitted disease foreseeable? the answer would surely have been yes. Rats are associated with disease but few people could specify which diseases are foreseeable'.

28 Such an approach has long been familiar in the context of harm that arises out of the breach of a statutory rule, see *Gorris v Scott* (1874) LR 9 Exch 125 and, more recently, *Empress Cars (Abertillery) Ltd v National Rivers Authority* [1999] 2 AC 22. There the court determines the scope of the duty (and hence the nature and extent of the harm for which the duty-breaker is liable) by reference to the legislature's intent in enacting the rule.

29 [1997] AC 191; sub nom *South Australia Asset Management Corp v York Montague Ltd* [1996] 3 All ER 365. See also *Reeves v Commissioner of Police for the Metropolis* [2000] 1 AC 360.

not liable for that portion of the claimant's loss attributable to the fall in the property market, Lord Hoffmann stated that, '[a] plaintiff who sues for breach of a duty imposed by the law (whether in contract or tort or under statute) must do more than prove that the defendant has failed to comply. He must show that the duty was owed to him and it was a duty in respect of the kind of loss which he has suffered.'³⁰

His Lordship drew an analogy with a doctor who negligently certifies a man with a bad knee fit to go mountaineering. He suggested that if the man were subsequently injured during the expedition, but in a manner that had nothing to do with the knee, a principle imposing liability on the doctor for such injuries would offend common sense. This is because 'it makes the doctor responsible for consequences which, though in general terms foreseeable, do not appear to have a sufficient causal connection with the subject matter of the duty'.³¹

However, this approach to defining the faulty risk is convincing only in a limited number of cases namely those where the parties have previously formed a relationship (often of a contractual nature). Thus in the *Banque Bruxelles* case, asking the question 'what risk of harm was the duty on the defendant imposed to guard against?' indeed provided an objective means of defining the risks for which the defendant was liable. That was because the nature of the contract between it and the claimant made it easy to infer the (limited) purpose of its duty to provide the proper valuation – in this instance to allow the claimant to retain a 'cushion' against (independent and uncontrollable) falls in the market.³² By contrast, in the majority of negligence cases (where there is no pre-existing relationship between the parties), asking the question, 'what is the purpose of the duty on the defendant to take care?' will merely elicit the response 'to eliminate or reduce the risks of injury to the claimant'. The issue of defining the risks encompassed within the duty is once again at large.³³

A recent case, involving no previous relationship between the parties, which illustrates this problem, is *Jolley v Sutton LBC*.³⁴ The facts were that the defendant council, through an administrative oversight, left a derelict boat lying on its land. The boat was accepted to be a risky allurement to children because of the chance they might injure themselves on its rotten planking. In the event the fifteen-year-old claimant and a friend tried to repair the boat, using a car-jack to raise it up, and the claimant suffered serious injuries when the boat toppled from the jack crushing him. The defendant argued that this injury was outside the risk that had made its failure to remove the boat faulty.

A strong Court of Appeal (Lord Woolf MR, Roch and Judge LJ) accepted this argument and held that the claimant's injuries were too remote.³⁵ However, it was reversed by a unanimous House of Lords, which took a wider view of the risks at issue and restored the trial judge's finding in favour of liability. Whatever the

30 *ibid* 211.

31 *ibid* 214.

32 Thus it is legitimate to distinguish, within the subsequent actual loss, between that which would have been absorbed by the cushion and that which would have been suffered in any event, only the former being recoverable.

33 A further criticism of the 'scope of the duty' approach is that encourages the conflation of the remoteness enquiry with the logically prior question of the duty of care. This results from inviting the judge to define the ambit of the duty owed in each case. In fact, the two enquiries have quite different functions. At the duty stage, a finding of no duty represents a determination by the court that the claimant's action falls within a category that, as a matter of policy, should not be recognised. At the remoteness stage, the general validity of the action is clear enough; the question is whether in the particular circumstances of the case the defendant should pay.

34 [2000] 1 WLR 1082.

35 [1998] 1 WLR 1546.

merits of their Lordships' decision,³⁶ it is apparent that here enquiring after the scope of the duty on the council to remove the boat provides no advance over the use of reasonable foreseeability alone. Indeed, this point was implicitly conceded by Lord Hoffmann, in his speech, where he advocated the merger of the two approaches in a large number of cases: 'the plaintiff must show that the injury which he suffered fell within the scope of the council's duty and ... in cases of physical injury, the scope of the duty is determined by whether or not the injury fell within a description which could be said to have been reasonably foreseeable'.³⁷

Ultimately, it is submitted that much of the difficulty in using the risk theory, and its tendency (by leaving so much to judicial discretion) to encourage a perception of the remoteness enquiry as entirely policy-driven, has stemmed from the failure properly to articulate the key concept of 'risk'. In particular, the concept has been too readily associated with the 'outcome harm' that results from negligent conduct. Instead, the question of whether the risk that made the defendant's conduct faulty subsequently materialised requires us to ask not so much 'what' that damage was, but 'how' it occurred. This involves focusing on the identity of the further elements in the causal set that completed for the harm in question. The proper understanding of the connection between risk and factual causation, which yields a revised version of the risk theory explanatory of the majority of remoteness problems, is pursued in the next section of this paper.

Causation and risk

Risk-claims and causal generalisations

When we claim that some conduct is 'risky', we appeal to our general experience that, over a range of previous instances in which there has been conduct of that type, harm has sometimes eventuated. Such a claim (hereafter a 'risk-claim'), which is often made before the effect of this particular conduct is known, takes a similar form to the causal counterfactual 'if "c", then "e".' However, whereas the latter asserts a relation of invariable sequence, a risk-claim merely expresses possible sequence: 'e' may occur, but then again it may not. In short, we are not committed to the proposition that the conduct will cause harm on *this* occasion.

Because 'risk' thus involves probability rather than certainty, it is sometimes thought that risk as a concept is *sui generis*, standing apart from causation (conceived of as invariable sequence).³⁸ In fact, though, the concepts are closely connected. This is because, as applied to everyday occurrences, the picture of causation embodied in 'if "c", then "e"' is a misleading one. Indeed this is so for two contrasting reasons. In the first place, it suggests that relationships of causal sufficiency (where 'e' always follows 'c') obtain between two single events in the world. But, as J.S. Mill saw, this is not the case. Instead, the 'c' term invariably denotes a complex of necessary conditions, which must be present for the effect to

³⁶ The case is discussed further at n 56 below.

³⁷ [2000] 1 WLR 1082, at 1091. The suggestion that it is the fact of *physical* injury having occurred that is critical here (as opposed to the lack of a pre-existing relationship between the parties) seems, with respect, mistaken, and indeed surprising given Lord Hoffmann's own previous invocation in *Banque Bruxelles* of the doctor-mountaineer hypothetical (which provides a counterexample).

³⁸ See eg Calabresi, n 2 above. Indeed, despite their illuminating discussion of how causal generalisations function as proxies for fully specified causal sets, Hart and Honore (n 3 above, 44 *et seq*) also fail to draw the connection, at this point, with the notion of risk.

follow.³⁹ Take, for example, the case of a person who drops a lighted cigarette into the wastepaper basket. Evidently, if the building in which this happens burns down, its destruction is not merely the upshot of the discarding of the cigarette: in addition, the presence of a host of other necessary conditions is required, both positive (eg inflammable material in the basket and oxygen in the air) and negative (eg the absence of a sprinkler system). It is only this complex of conditions, a 'causal set', that, taken as a whole, is sufficient for the outcome.⁴⁰

Secondly, though (and by way of explaining the tendency to conceive of 'c' as a single event), in citing something as the cause of something else, we do not in practice appeal to the laws of invariable sequence that pertain between a full causal set and its consequent. Rather, in ordinary discourse, we rely on causal generalisations, which take as their starting point some human conduct or other unusual condition and chart the impact of such interventions upon a neutral background environment (ie one assumed to consist of more or less standard further conditions). As Hart and Honoré write:

When we assert that A's blow made B's nose bleed or A's exposure of the wax to the flame caused it to melt, the general knowledge used here is knowledge of the familiar way to produce, by manipulating things, certain types of change which do not normally occur without our intervention. If formulated they are broadly framed generalisations, more like recipes, in which we assert that doing one thing will 'under normal conditions' produce another, than statements of 'invariable sequence' between a complex set of specified conditions and an event of the given kind.⁴¹

As they note, such generalisations are well adapted for use in an inquiry *ex post facto* as to why some untoward outcome has occurred. At this moment in time, what we are after is an explanation of the outcome, and this involves the search for a factor (often faulty human agency) that sets this case apart from the general run of cases where things go on as normal. It is this point that explains why, in the example of the building burning down, we would normally cite the discarding of the cigarette as 'the cause'. By contrast, the other coexisting factors, whose presence was also required, will be relegated to the status of 'mere conditions':

In most cases a person who is told that there has been a railway accident in which lives have been lost, or a house that has burnt down, and who is moved by this to ask what was the cause of the accident or fire, would know of or assume the existence of many factors. These are treated as mere conditions because they are present both in the case of the disaster and of normal functioning.⁴²

On the other hand, the fact that causal generalisations stand as a proxies for the full causal sets sufficient for any given outcome has significant implications *ex ante* when we are planning to introduce some *new* change in the world. In particular, given our intervention alone, we cannot be certain that the outcome will occur: on this occasion some of the further conditions required to complete

39 J.S. Mill, *System of Logic*, Book III, Chapter 5.

40 It is this fuller understanding of factual causation that informs the NESS test (improving upon the 'but for' test in causal over-determination cases). According to this test, an event will qualify as a factual cause of another event to the extent that it was a necessary element in a causal set sufficient to produce the latter: see Wright, n 7 above.

41 Hart and Honoré, n 3 above, 31.

42 *Ibid* 34. The distinction between 'a cause' and the further standard necessary conditions is not intended to be a hard and fast one, and will vary both according to what is 'normal' in the surrounding environment, and the particular concerns of the investigator, see *ibid* 35-37, and J. Feinberg, *Harm to Others* (New York: OUP, 1984) 176-177. For the purposes of this paper, a condition will be treated as a cause in so far as, in the *ex post facto* legal inquiry, it would be mentioned in any reasonably complete account of how a given accident occurred.

the causal set may be absent. Nevertheless, as Hart and Honore's comparison of them to 'recipes' suggests, causal generalisations here retain considerable utility. This is because they incorporate an implicit appeal, based on our broad experience of similar previous cases, both as to the kind of further conditions required to complete a causal set and, how often they are present in the surrounding environment. Thus, while we cannot predict for certain that the outcome will occur, our confidence will be justifiably high if, in nearly all such cases, it has done so: in effect such a finding tells us that the further conditions almost always appear.

Consider, for example, a case where a farmer plants some seeds in order to grow a crop of potatoes: it of course remains possible that one or more of the various further conditions required, going to such matters as the suitability of the soil and climate, will be absent – perhaps there will be a drought that year. Equally, some of the crop will be damaged by birds and insects. (Indeed more outlandish possibilities cannot be excluded: a leak from the local factory may pollute the land, or a rival farmer come and dig the crop up, etc.) Even so, from his experience of growing potatoes over the years, the farmer will have a good idea of how often such contingencies tend to arise and, based upon this, be able to estimate the likely crop yield this year.

In the above case, the outcome was one whose production was desired. However, the same conception of causal generalisations as 'recipes' also provides a useful approach to understanding risk-claims. In particular, where some change in the world threatens to be of a harmful character, such a claim may be likened to a *negative recipe*. Suppose for example that, in the case mentioned earlier of the building destroyed by fire, we were present just as the person was about to throw their lighted cigarette into the waste-paper basket; our warning 'Watch out' is shorthand for 'General experience has shown that "r" (the discarding of a lighted cigarette) is, given the presence of certain further conditions, followed by "e" (a serious fire). Therefore if you do not want to court the possibility of "e", you should not do "r".'

Here, the risk-claim is framed by reference to the potential presence of further conditions, which, in an *ex post facto* enquiry as to what had caused the fire, would normally escape comment: they are 'mere conditions'. However, it is important to note that, on other occasions, a claim may be predicated upon the possible completion of a causal set for harm in which the appearance of some further condition of a non-standard character is also required. In that case, if the harm materialises, the further condition too has a claim to causal status. Thus if, during breakfast, I put my cup of coffee down on the edge of the table, this will be called risky because, given some further slight clumsiness on my part or that of my neighbour – familiar enough from experience – the cup will be knocked from the table. Here, in a description of the accident after the event, the later act of clumsiness will also be cited as a 'cause' of it.

Moreover, this last example illustrates another highly significant feature of risk-claims. In particular, by leaving matters relatively vague as to which further conditions may appear, they often allow for different causal sets to complete for the harm. Here for instance, the accident may occur either through my further clumsiness (the complete causal set will contain my earlier risky conduct and *my* subsequent act, plus further standard conditions) or my neighbour's (the complete causal set will contain my earlier risky conduct and *his* subsequent act, plus the standard conditions). Indeed it is true of almost all faulty conduct that, at the moment it takes place, not only may no harm at all result (if no causal set completes around it) but, insofar as it does, this could occur through its presence in

alternative causal sets. This 'pleni-potentiality' of risk is of importance when we come, *ex post facto*, to consider if the harm that has materialised is too remote.

The enquiry *ex post facto*

In the negligence enquiry we are concerned to attribute responsibility for harm that has already occurred. *Ex hypothesi*, by the time that we reach the question of remoteness, a full causal set including the defendant's faulty conduct (ie conduct that posed an unreasonable risk of harm to the protected interests of another) has completed for the harm. What is now required is a comparison of that causal set with the causal generalisation/risk-claim in virtue of which the conduct, regarded *ex ante*, counted as faulty. In short, do the two match? Answering this requires us to pay close attention to the further necessary conditions (besides the faulty conduct) that completed the set.

In cases where the faulty conduct of the defendant ranks, *ex post facto*, as the sole 'cause' of the event (ie the further conditions that complete the set are mere conditions) there will be no problem. This is because the defendant, as a reasonable person, must be taken to have foreseen as a matter of course the standard features of the background environment in which he operated. Indeed, in this regard, the precise detail of such conditions, and the possibility of their having combined in a statistically infrequent manner, is irrelevant. Thus consider an example in which a faultily secured chandelier, comprising part of the scenery in a play, falls onto the stage during a performance and shatters. None of the cast is hurt, but a shard of glass flies into the audience and pierces the heart of a man in the third row, killing him instantly. Here, though highly unfortunate, the result, in being produced by various standard conditions combining with the risk factor provided by the defendant, straightforwardly substantiates the relevant risk claim. This can be tested by asking if, *ex post facto*, there is anything we need to know about the way the injury occurred apart from the defendant's faulty conduct: the answer is no.⁴³

By contrast, difficulties in remoteness begin in those cases where the harm that has occurred involves some further 'causal condition' featuring in the set for the harm. Of course, as we have seen, sometimes the risk-claim attaching to defendant's conduct will have been based precisely on the possibility (and necessity) of such a condition appearing and, clearly, here its subsequent appearance cannot take the damage 'outside the risk'. Thus in the case of *Pearson v Lightning*⁴⁴ the claimant, who was standing on an adjoining fairway, was struck in the eye by the defendant's golf ball after it ricocheted off a bush. The defendant, in attempting a difficult chip over the bush, had failed, contrary to normal practice, to wait for the claimant to move out of range before striking his shot. Here, in so far as the bush would be mentioned in an account of the accident, it was precisely the danger of a ricochet from it that, in the circumstances, made the defendant's conduct faulty.⁴⁵

43 As noted above (see text accompanying n 15), in negligence, liability for harmful changes in the world is mediated through the foresight of a reasonable person, and such a person must be taken to have foreseen as a matter of course the 'more conditions' that pervade the background environment. It is submitted that this is the true principle lying behind cases such as *Vacwell Engineering v BDH Chemicals* [1971] 1 QB 88, in which recovery has been permitted notwithstanding that the detail and extent of the harm was unforeseeable.

44 (1998) *The Times*, 30 April 1998

45 See also *Stansbie v Troman* [1948] 2 KB 48, discussed at n 95 below. It may be noted, in passing, that unlike the case with 'outcome harm', there is little scope for a court to vary the description of the further causal condition in the set, eg by suggesting that while a ricochet off a bush was reasonably foreseeable, one off a juniper bush was not. To the extent that this had no bearing on the accident (ie it was unnecessary for the harm) the detail as to the species of bush will be dismissed straightaway as irrelevant.

In many other cases, however, the risk-claim will have been founded in the first instance upon the propensity of the risk factor alone, in combination with mere conditions, to produce the harm, ie this already sufficed to make the conduct faulty. And here there will be the potential for mismatches to occur between the prior claim and the causal set in suit. In particular, it must now be asked if the potential completion of *that* set (including, as it does, the further causal condition) would itself have deterred a reasonable person from the conduct that was the subject of the risk-claim.

Reasonable foreseeability again

To answer the last question affirmatively, requires, as a minimum that there was a known risk of the conduct completing a causal set by combining with the other causal factor. However, the defendant's faulty conduct will only amount to a risk in this respect if there is a general link *over a range* between it and the harm. More specifically, experience must show that a set sufficient for the harm, including both *the conduct* and the other causal condition as members, appears more often than a set including the latter alone, – if not, the harm will count simply as a 'coincidence'. In this context, Hart and Honore discuss a hypothetical case in which a coachman falls asleep at the reins of his coach.⁴⁶ The coach deviates from its proper route and the passenger is injured when it is struck by lightning. Here the risk-claim in virtue of which the defendant's falling asleep was faulty (ie that the coach might run off the road or collide with other traffic) has not been substantiated. Instead such conduct has figured quite fortuitously in a causal set for the lightning damage: a causal set for the damage was just as likely to have completed in its absence. In this connection McHugh J commented in *Chappel v Hart* that '[I]f the act or omission of the defendant has done no more than expose the plaintiff to a class of risk to which the plaintiff would have been exposed irrespective of the defendant's act or omission, the law of torts should not require the defendant to pay damages'.⁴⁷

Secondly, even where a general link between the faulty conduct and the harm is apparent, the presence of the further causal condition in the background environment, and its potential to combine with the faulty conduct in a causal set for harm, must have been known of at the time of the conduct. A good illustration of a case where such knowledge was absent is *Doughty v Turner Manufacturing Ltd*.⁴⁸ Here one of the defendant's employees knocked an asbestos lid into a vat of molten liquid, an act which was faulty because of the risk that other employees might be splashed. No splash occurred but moments later, due to a chemical reaction (whose existence had not previously been suspected), there was an eruption of the liquid in which the plaintiff employee was injured. The Court of Appeal held (it is suggested correctly) that the damage was too remote.⁴⁹

In the last two scenarios, the defendant's faulty conduct had no, or no known, propensity to feature in the causal set that actually materialised for harm. Thus, in respect of *that possibility*, it would not have been the subject of a hypothetical risk-claim at all. Nevertheless, on many other occasions experience will recognise that the defendant's conduct did increase the chance of the causal set in suit completing. Here, a risk-claim might have been uttered, at the moment of the

46 See n 3 above, 470–471. As they note, the hypothetical was originally proposed by the nineteenth century German jurist, von Kries.

47 (1998) 72 ALJR 1344, at 1350.

48 [1964] 1 QB 518.

49 See also *Roe v Minister of Health* [1954] 2 QB 66 and *The Wagon Mound (No 1)* [1961] AC 388.

defendant's conduct, predicated upon the possibility of *that set*. The next question, though, is whether this risk-claim was one that, in the circumstances, a reasonable person would have heeded.⁵⁰

In this regard, the materialisation of a causal set for harm, which contains as necessary elements both the defendant's faulty conduct and the further causal condition, must not be 'far-fetched'. Thus, consider a variant of *Pearson v Lightning*⁵¹ in which the golf ball would have passed the claimant harmlessly by if it had not been deflected by striking a bird in mid-air. Certainly the *theoretical* possibility of such a thing happening is recognised (it conforms to our knowledge of causal laws in the world), but even where the costs of guarding against it are negligible, a reasonable person would not do so. The possibility of such a combination of two events, whose probabilities taken singly are already so small, will simply be dismissed.⁵²

On other occasions, however, the harm, while it does not substantiate the principal risk-claim that would have been made at the time of the conduct, will involve the materialisation (through the conduct combining with a further causal condition) of a known *subsidiary* risk, and one that is not far-fetched. It is at this point that the source of the conduct's faultiness will need to be reappraised. In particular, leaving aside the principal risk (ie the chief reason for regarding the conduct as faulty), would the possibility of the subsidiary risk itself have provided sufficient reason for a reasonable person not to engage in the conduct?

Suppose, for example, that we are told that a building has burnt down after the defendant accidentally discharged a round from his gun into the ceiling and our enquiries show that the fire began after the bullet cut through an electrical cable. Here, the principal risk-claim made in respect of the defendant's conduct would not have adverted to the possibility of a fire resulting (we would instead have first thought in terms of the bullet killing or injuring someone in the room above or causing property damage by shattering or holing objects). Nevertheless, given that the presence of the cable as a further causal condition leading to the fire is by no means far-fetched, in such a case (where there was no valid reason to fire the gun) liability will readily extend to harm arising from this subsidiary risk. A case involving the presence of such a reasonably foreseeable subsidiary risk is that of *The Wagon Mound (No 2)*.⁵³ Here, in contrast to the first *Wagon Mound* case,⁵⁴ it was found to be theoretically foreseeable that bunkering-oil floating on water could catch fire, albeit the risk remained very small. However, given that the discharge of the oil was entirely unjustified on the facts, the defendant was liable for the fire that followed.

By contrast, there will be other cases where the possibility of a causal set for harm completing, featuring the further causal condition, was sufficiently unlikely, and the opportunity cost of foregoing the conduct sufficiently great, for it to have been reasonable to run *that* risk. Thus suppose a variant of *The Wagon Mound (No 2)* in which there was some good reason to discharge the oil – eg the ship was believed to be damaged beneath the waterline and her crew wished to lighten her in order to carry out an inspection. It may be that such conduct would remain faulty relative to the risk of fouling but not relative to the much less likely one of fire.⁵⁵

50 In short, though theoretically foreseeable, was the risk one that was *reasonably* foreseeable?

51 See n 44 above.

52 Note that this is not a 'coincidence' case: the probability of the outcome in question is greater (indeed infinitely so) given the presence of the defendant's faulty conduct, as opposed to in its absence.

53 *Overseas Tankship (UK) Ltd v The Miller Steamship Co Pty Ltd, The Wagon Mound (No 2)* [1967] 1 AC 617.

54 The background to the different finding of fact in the two cases is explained at n 20 above.

55 In this regard, the fear of the wharf-owner in *The Wagon Mound (No 1)* of being non-suited due to contributory negligence may have been unjustified given the utility in continuing the welding works.

The analysis applied: *Jolley v London Borough of Sutton*

In the light of the above analysis, the remoteness issues that recently arose in *Jolley v Sutton LBC*⁵⁶ will now be considered in more detail. As noted earlier, the case arose out of a breach of duty by the council in failing to remove a derelict boat lying on its land. The claimant suffered serious injuries when the boat, which he and a friend had jacked up to try and repair, fell on top of him.

The trial judge took a wide view of the risk presented by the defendant's breach, finding that, 'the type of accident and injury which occurred ... was reasonably foreseeable (albeit that it involved significant meddling with the boat by two young teenage boys and that the injuries proved to be very severe)'.⁵⁷ However, in reversing this judgment, the Court of Appeal took a narrower view of matters. In its view, what had made the council's failure to remove the boat faulty was the risk that, given the boat's rotten condition, children playing on it might injure themselves if its planking gave way. The claimant's injuries, on the other hand, did not stem from the rotten condition. In the words of Lord Woolf MR:

It was a combination of ... two features that made it the duty of the Council to have the boat removed ... However these features, the attractiveness of the boat to children and its dangerous condition, were not established to be part of the causes of the accident. The immediate cause of the accident was that the two boys jacked and propped the boat up so that they could work underneath it and did so in a way that meant the boat could and did fall on the plaintiff.⁵⁸

Accordingly, this accident was 'of a different type and kind from anything that [the defendant] could have foreseen'.⁵⁹

This judgment was in turn reversed by the House of Lords, which restored the decision of the trial judge. It did so by the simple expedient of finding that the latter's initial judgment, that the harm was reasonably foreseeable, was one the Court of Appeal had not been entitled to disturb. In the words of Lord Steyn:

The issue whether an accident of the particular type was reasonably foreseeable is technically a secondary fact but perhaps it is more illuminating to call it an informed opinion by the judge in the light of all the circumstances of the case. In my view it was an opinion which is justified by the particular circumstances of the case. Counsel has not persuaded me that the judge's view was wrong.⁶⁰

Nevertheless, while on the facts the outcome in favour of liability may be supported (see below), it is submitted, with respect, that the House of Lords' approach in *Jolley* is one that owes more to pragmatism than principle.⁶¹ Instead it is the Court of Appeal's judgment, which, though ultimately flawed, shows greater promise in analytical terms. Of particular interest was the attempt by Lord Woolf MR to contrast the instant facts with those of *Hughes v Lord Advocate*⁶² in which another child claimant suffered injury after a paraffin lamp, which he knocked into an open manhole unexpectedly exploded. In *Jolley* Lord Woolf MR attached particular weight to the following passage in Lord Reid's speech:

⁵⁶ [2000] 1 WLR 1082.

⁵⁷ [1998] 1 Lloyd's Rep 433, 440. The judge discounted the damages he awarded by 25 per cent to reflect the claimant's contributory negligence.

⁵⁸ [1998] 1 WLR 1546, 1553.

⁵⁹ *ibid* at 1554.

⁶⁰ [2000] 1 WLR 1082, 1089.

⁶¹ It is interesting to speculate on what the House of Lords' approach might have been had the trial judge in *Jolley* found the harm in suit *unforeseeable*, and the Court of Appeal's reversal of this been the subject of an appeal by the council.

⁶² [1963] AC 837.

The ground on which this case has been decided against the appellant [in the lower courts] is that the accident was of an unforeseeable type. Of course the pursuer has to prove that the defendant's fault caused the accident, and there could be a case where the intrusion of a new and unexpected factor could be regarded as the cause of the accident rather than the fault of the defender. But that is not this case. The cause of this accident was a known source of danger, the lamp, but it behaved in an unpredictable way.⁶³

Lord Woolf MR went on to comment that,

Lord Reid later by way of distinction referred to a situation where there was an intrusion of some new and unforeseeable cause like the falling of a ceiling so that the damage cannot be said to have resulted from the defendant's breach of duty and in relation to that situation he appears to accept that there would be no liability.⁶⁴

It is apparent, that the Master of the Rolls, in denying liability in *Jolley*, must have felt that the latter case was in some way akin to the one of the falling ceiling.⁶⁵ However can such a distinction between its facts, and those of *Hughes*, be sustained?

In terms of the analysis suggested in this paper, it is apparent that harm in *Hughes* could have resulted from causal sets forming either around the manhole or around the lamp, each such risk providing an independent reason for regarding the defendant's conduct as faulty.⁶⁶ In the event, the causal set that completed for the harm included *both* manhole and lamp as members together with those further conditions. Here, even if such an upshot was relatively improbable, the defendant, having faultily supplied both such causal conditions, could not appeal to the unlikelihood of their combination as a basis for excusing him from liability.⁶⁷

Certainly, in this regard, *Jolley* does differ at first sight from *Hughes*. In particular here it was the claimant himself, who by jacking the boat up, supplied the further causal condition that, together with the defendant's faulty conduct in leaving the boat on its land, completed the causal set for harm. However, rather than treating this fact as 'the immediate cause' of the accident (an obscure and redundant term), the Court of Appeal should at this point have asked if a reasonable person would have foreseen the possibility of a causal set completing in which, besides the defendant's failure to remove the boat, such an act figured.

Now it is arguable in this regard that, had the claimant turned up with the jack 'out of the blue' and suffered his injuries more or less straightaway, this would not have been reasonably foreseeable. Such a possibility, while *theoretically* foreseeable (and not far-fetched) would, in the light of the costs of removing the boat, not have moved a reasonable landowner to act. However, what makes the case, by contrast, quite straightforward on its facts is that, at the time of his accident, the claimant had already been using the jack to prop up the boat for several weeks. In this context, the defendant may be said to have had constructive

⁶³ *ibid* 845.

⁶⁴ [1998] 1 WLR 1546, at 1552.

⁶⁵ The falling ceiling analogy, suggestive, as it is, of coincidental injury, is on the face of things not at all persuasive. After all in *Jolley*, far from being a coincidence, the defendant's failure to remove the boat created the risk of injury to the plaintiff. However, in so far as the analogy is taken to refer more broadly to the possible intrusion into the causal set of some further causal condition that (while rarely a coincidence) was not reasonably foreseeable, its invocation is more plausible.

⁶⁶ There were two concurrent risk-claims linking either the open manhole (plus the presence of an inquisitive child and further mere conditions) or the unguarded paraffin lamp (plus the same further conditions) to harm, and the defendant's conduct was faulty in respect of each.

⁶⁷ The same analysis can be used to support the result reached in *Re Polemis*. In particular, not only was the charterer there responsible for the acts of the stevedores who dropped the plank, but also (it may be assumed) for the previous escape of the petrol vapour which had, in effect, turned the ship into a floating bomb.

knowledge of the latter's potential to figure as a further causal condition in the set.⁶⁸ This being so, the case does once again assume a closer similarity to *Hughes*, in which the defendant had notice of both the causal conditions which, together with mere conditions, completed the harmful causal set. It follows, equally, that the ultimate finding in favour of the claimant in the House of Lords was justified.

'Policy' exceptions

The model proposed in the previous section represents, it is suggested, the default position. The defendant must have unreasonably created or added to the risk of the harm in suit in the sense that its occurrence substantiates a risk-claim that (hypothetically) would have been made at the time he acted. The model tallies with the idea of persons as purposive agents intervening in the world in pursuit of their own ends, while not exposing others to unreasonable risks of harm. Where the defendant's conduct does play a necessary part in some harm, but the possibility of the relevant causal set completing (which includes some further causal condition) would not have deterred a reasonable person from acting, the harm is not attributable to the agency of the defendant.

Nevertheless, in practice there are a number of exceptions to this position, where 'policy' in a restricted sense⁶⁹ will operate to modify liability. In the first place there are certain categories of case where, though the defendant did not negligently create or add to the risk of harm that actually transpired, he will still be held liable. In effect, the risk of a causal set completing, which includes a further causal condition that was not reasonably foreseeable, is shifted onto the defendant. Secondly, and conversely, there are also some cases where the defendant will be excused liability for harm even though his conduct was a necessary condition in a causal set, the completion of which was reasonably foreseeable.

The defendant is liable although the presence of the further causal condition in the set was not reasonably foreseeable

The two main contexts in which the defendant's liability will extend beyond the reasonably foreseeable risk posed by his conduct comprise, first, the 'eggshell skull' rule and, secondly, certain cases of 'ulterior harm'.

According to the former rule, provided the claimant has suffered some reasonably foreseeable personal injury, the defendant is liable for the whole injury sustained, even though, due to some unusual susceptibility of the claimant, its full extent was not reasonably foreseeable. The working of the rule is exemplified by the House of Lords' decision in *Smith v Leech Brain*,⁷⁰ in which the claimant, through his employer's breach of duty, suffered a burnt lip. Unfortunately, because of a latent disposition of the claimant, the wound became cancerous and he died. His widow was entitled to recover damages in respect of the death.

68 It could be argued that, in the light of its primary failure to remove the boat, the Council came under a secondary duty at least to keep an eye on things. Accordingly, knowledge of the claimant's activity with the jack could readily be imputed to it.

69 The considerations at issue remain situation-specific in a way that the policy arguments that inform the duty of care question (involving broader questions of distributive justice) generally do not. It is for this reason that 'policy' has been placed in quotes.

70 [1962] 2 QB 405. See also *Warren v Scrutton's Ltd* [1962] 1 Lloyd's Rep 497.

To put the eggshell skull rule into the terms of a causal set analysis, the defendant has contributed a risk factor, which is faulty relative to the possible appearance of certain other conditions completing a full set for harm. Those conditions have indeed appeared, but so too has a further causal condition going to some idiosyncrasy in the claimant's physical makeup, which has had the effect of worsening the harm. The rule, in effect, says that, once the set containing the anticipated conditions has materialised, resulting in actionable damage, the presence in the set *in addition* of the further, unforeseeable condition is beside the point.⁷¹ The risk of it appearing is simply placed on the defendant. The rule is often justified on the basis adverted to by Glanville Williams, when he wrote in 1961: '... when the plaintiff has suffered bodily injury, it is difficult to maintain the cold logical analysis of the situation. Human bodies are too fragile, and life too precarious, to permit a defendant nicely to calculate how much injury he may inflict without causing more serious injury'.⁷² However, another consideration is that the claimant cannot avoid carrying his weakness around with him it is part of him, and a rule requiring him to desist, on this account, from social intercourse would be both impracticable and oppressive. Admittedly, an eggshell skull claimant cannot complain if he submits himself to a situation in which there is a risk of *inadvertent* injury, and is so injured. As Lord Wright commented in *Bourhill v Young*, 'One who suffers from the terrible tendency to bleed on slight contact ... cannot complain if he mixes with the crowd and suffers severely, perhaps fatally, from being merely brushed against. There is no wrong done there ...'.⁷³ But an eggshell skull claimant is protected from injury to the extent that, in the same circumstances, a normally constituted person would have suffered injury of an actionable nature (albeit a lesser one).

Besides the limited exception to the risk theory embodied in the eggshell skull rule, a more difficult problem for the theory has sometimes been thought to arise from cases of 'ulterior harm'. These occur when reasonably foreseeable initial harm subsequently combines with some new condition, producing further, secondary harm.⁷⁴ In discussing such cases, we must be clear, from the outset, as to what amounts here to *further* harm (as distinct from initial harm that simply develops over time while remaining 'the same'). It is suggested that two factors are critical. First, the claimant must suffer some new and identifiable setback to his interests.⁷⁵ Secondly, it must be possible, in theory, for some agent to have intervened and prevented the occurrence of the new setback (though the failure to do so will not be negligent).⁷⁶

Hart and Honoré argue that ulterior harm cannot by definition be part of the harm the risk of which was our reason for calling the defendant's conduct faulty:

71 Here, significantly, *some* injury to the claimant would have occurred even in the absence of that further condition appearing.

72 'The Risk Principle' (1961) 77 LQR 179, 196.

73 [1943] AC 92, 109.

74 See Hart and Honoré, n 3 above, 263.

75 See Feinberg, n 42 above, Chapter 1. The failure to make clear the need for a new set-back to the claimant's interests has sometimes resulted in claims that properly go to *quantum* being dealt with as ulterior harm problems in remoteness – see eg the Court of Appeal's decision in *The Sivad* [1998] 2 Lloyd's Rep 97 in which it is suggested, with respect, that the minority reasoning of Hobhouse LJ (as he then was) is to be preferred.

76 See J. King, 'Causation, Valuation and Chance in Personal Injury Torts Involving Pre-existing Conditions and Future Consequences' (1980–1) 90 Yale LJ 1353. It follows that when identifiable harm has occurred, which includes a portion of latent damage that is certain to manifest itself later to the claimant's further detriment (as happened in *Smith v Leech Brain*), we do not classify the latter as 'ulterior'.

[W]here the risk of harm is created by a situation, the risk of which occurring was itself enough to deter a reasonable man, to ask whether he would be deterred by the further harm alone is to put a question he cannot answer. To press it would be to ask the reasonable man first to imagine himself in a situation the risk of which would have deterred him from acting, and then to tell us whether in view of the further risk thus created he would have been deterred.⁷⁷

In terms of the analysis offered in this paper, this amounts to the proposition that a reasonable person will not look beyond the possible completion of an initial causal set for harm around a given piece of conduct. In so far as the conduct is faulty relative to *that* possible harm, this already provides a sufficient reason not to engage in it; the possibility that the harm may in turn feature as a necessary condition in a secondary causal set, producing further harm, is left out of account. Hart and Honoré then proceed to use the undeniable phenomenon of recovery for such latter harm being permitted in some instances as a major plank of their argument against a 'generalised risk theory' to explain remoteness cases.⁷⁸

However, it is suggested, with respect, that they make too much of such cases. In the first place, it is important to distinguish the relation between the defendant's initial act and harm caused to multiple distinct claimants. There is, in particular no reason why harm that has befallen one claimant should not simultaneously pose a risk to a second claimant of new harm, and (even though it presupposes the harm to first claimant) there seems to be nothing artificial in asking if, at the time the defendant acted, harm to the second claimant was reasonably foreseeable. Thus suppose X carelessly shunts Y's car off a railway bridge onto the track below. In determining whether X is also liable for injuries to Z, a passenger on a train which, moments later, derails on hitting Y's car, the use of reasonable foresight is perfectly coherent. This represents an appeal to the circumspection of the reasonable person *as he is driving over a railway bridge*, not to that of the unreasonable person who has already shunted a car onto the track. One of the key cases used by Hart and Honoré to support their argument is of this form, namely *Re Guardian Casualty Co.*⁷⁹ Here the defendant's negligent driving caused him to mount the kerb and crash into a building. Some twenty minutes later, while the car was being pulled clear by firemen, a stone from the building's cornice, which had been loosened by the impact, fell onto the pavement nearby, killing the claimant. However, while it is suggested that it was open in principle to have found that the claimant's injury was reasonably foreseeable, it is further suggested that the court arrived at the wrong result on the facts: a reasonable person would not have been deterred by such a (far-fetched) risk at the time of driving.

Secondly, whilst Hart and Honoré's insight holds good for cases where the *same* claimant suffers two consecutive injuries (the first being necessary for the second),⁸⁰ the recovery that is sometimes allowed can, as with the eggshell skull rule, be explained as a special departure from the risk theory. In particular, in the ulterior harm context, the common feature linking the relevant cases is that the initial injury suffered (for which the defendant is undoubtedly liable) rendered the claimant *vulnerable* to such further harm.⁸¹

⁷⁷ See Hart and Honoré, n 3 above, 264

⁷⁸ *ibid* 263 *et seq.*

⁷⁹ 253 App Div 360 (1938). For other cases of this type, where liability towards an ulterior claimant has been denied, see *Palsgraf v Long Island RR Co*, 248 NY 339, 162 NE 99 (1928), *Crankshaw v Piedmont Driving Club*, 115 Ga App 820, 156 SE 2d 208 (1967), and *Wright v Lodge* [1993] 4 All ER 299.

⁸⁰ eg Y himself, in the railway bridge example, suffers further injury when the train hits his car.

⁸¹ See D. Payne, 'Foresight and Remoteness of Damage in Negligence' (1962) 25 MLR 1, 12.

In fact such 'vulnerable claimant' cases may take a number of forms. Sometimes the first injury will physically weaken the claimant and lower his normal resistance to some further external source of injury, as when a wound becomes infected. It is apparent that, at least where there is no negligence (either contributory or on the part of those treating the claimant), full recovery for the further injury will be allowed.⁸² Indeed the defendant bears the risk that the medical intervention required to treat the initial injury may actually *add to the harm*.⁸³ Thus in *Robinson v Post Office*, the claimant suffered a scratch at work through the defendant's negligence and was taken to hospital for a tetanus jab. As a result of suffering a serious reaction to the jab, he was subsequently left brain-damaged. The defendant was held liable in full.⁸⁴ *Robinson* is notable for involving a combination of the vulnerable claimant and eggshell skull rules. Thus the risk was placed on the defendant both as regards the general possibility that treatment required to counteract its earlier wrong would further damage the claimant *and* the possibility that a particular quirk in the plaintiff's metabolism would contribute to such damage. On the other hand, it was conceded (it is suggested rightly) by counsel in the case that, had the doctor's negligent failure to perform an allergy test been causative of the brain damage, this would have relieved the defendant of liability for the latter damage.

In the second place there are cases where, for a time, the original injury serves to deprive the claimant of his autonomy and, in that state, he is unable to protect himself from further injury. This was the case where X shunted Y and his car onto the railway track. Indeed, assuming Y is trapped in the car, it is suggested that this fact alone (ie the deprivation of his freedom of movement) would be sufficient to call the principle into play even if he were as yet physically unscathed. It is submitted that this principle explains the finding in favour of liability in the case of *Versic v Connors*,⁸⁵ where the claimant's husband was a passenger in a car that overturned following a collision caused by the defendant's negligence. His head acted as a dam to a gutter at the side of the road and he drowned in a few inches of water (a risk that otherwise would have been dismissed as far-fetched). The same principle applies *a fortiori* to those cases in which the earlier injury has rendered the claimant insensible. Thus in America defendants, whose negligence knocks the claimant unconscious, have been held liable for opportunistic thefts from the latter by third parties.⁸⁶

Finally there are cases where the claimant does not lose all autonomy, but retains some choice of action, and how he conducts himself then determines whether he suffers further injury. In such cases of reduced autonomy, the courts will allow claimants a 'period of grace' to adjust to their new circumstances (involving the

82 *Keegan v Minneapolis & St Louis Rly Co*, 76 Minn 90 (1899); *Adelaide Chemical & Fertiliser Co Ltd v Carlyle* (1940) 64 CLR 514. Such cases are very close to the 'eggshell skull rule' except that, instead of being part of the claimant's physical being, the further causal condition was external to him.

83 Arguably, though, the possibility of the further harm from the treatment should at least have been *theoretically foreseeable*; the defendant is not required to subsidise new discoveries in medical science. Thus in the South African case of *Alston v Marine & Trade Insurance* (1964) (4) SA 112 (W) the defendant was not liable when the claimant, who required medication as a result of the injury he inflicted, suffered a stroke after the medicine reacted (in a hitherto unsuspected way) with the cheese sandwich he was eating.

84 [1974] 2 All ER 737; see also n 8 above.

85 [1969] 1 NSW 481 (New South Wales CA).

86 *Brauer v New York Central & HRR Co*, 91 NJL 190, 103 Atl 166 (1918). Here, however, our inclination to hold the defendant liable may be tempered by the countervailing consideration that the further loss resulted from deliberate third party agency. This question is looked at further below (see n 102 *et seq*).

diminution of their previous capacities by the defendant's negligence), and during which time the risk of further injury remains with the defendant.⁸⁷ Similarly, where the initial injury leaves the claimant in a perilous position, where he must actively struggle to avoid further damage, and he non-negligently acts in a sub-optimal way, the risk of the further harm will be placed on the defendant.⁸⁸

In all of the above cases, importantly, the unreasonable risk to the claimant's interests, which the defendant's faulty conduct presented, has materialised. In respect of the initial harm caused, there is no doubt as to the defendant's agency. Where that harm, as one of its effects, specifically renders the claimant vulnerable to further harm, it seems right that, if such harm occurs, the defendant should pay.⁸⁹ Nevertheless, such cases apart, it is submitted that the general tenor of the courts' decisions (in keeping with the risk theory) has been to reject recovery for ulterior harm. The presumption is that harms that have already happened are simply part of the background world with which reasonable persons must contend. Accordingly, the onus is on the latter to take the necessary steps to prevent such past harms from giving rise to further injury. This is probably the best way of understanding the result reached by the Court of Appeal in the well-known case of *Lamb v Camden LBC*,⁹⁰ where the claimant, whose house was negligently damaged by the council, failed to recover in respect of further damage done by squatters while the house lay empty. Though in this case third parties inflicted the damage in question, it is suggested that the same result would apply if the house, while empty, had suffered natural dilapidation – eg a colony of mice had moved in and eaten the remaining fixtures.

The defendant is not liable although the presence of the further causal condition in the set was reasonably foreseeable

As previously noted,⁹¹ as well as limiting liability in cases where an application of the *Polemis* rule would have found it (ie where 'direct' harm was not reasonably foreseeable), the risk theory/reasonable foreseeability test may sometimes operate to extend it. This will occur where harm, which, on the older approach would have been classed as 'indirect', is nevertheless reasonably foreseeable. Generally, the courts have been reluctant to accept this 'extensive' aspect of the risk theory and this explains their continued appeal, in some cases, to that mainstay of the 'direct consequences' approach, the *novus actus interveniens* that 'breaks the chain of causation' between the defendant's faulty conduct and the harm.⁹² In fact, though,

87 See *Wieland v Cyril Lord Carpets* [1969] 3 All ER 1006. Sometimes, however, the claimant's behaviour will be regarded as sufficiently autonomous to remove the risk from the defendant in relation to the further harm, see *McKew v Holland, Hannen & Cubitts* [1969] 3 All ER 1621, discussed at n 99 below.

88 *Lord v Pacific Steam Navigation Co Ltd, The Oropesa* [1943] P 32. See also *The City of Lincoln* (1889) 15 PD 15. Similarly, if a third party intervenes as a 'rescuer', to try and counteract the perilous situation created by the defendant (both prior to any harm having occurred, and subsequently, to mitigate its effects), the courts have generally placed the risk of injury to the rescuer on the defendant, see *Haynes v Harwood* [1935] 1 KB 146; *Baker v TE Hopkins* [1959] 1 WLR 966; *Ogwo v Taylor* [1988] AC 431.

89 Indeed it could be argued that the initial harm includes the compromised resistance to the further harm.

90 [1981] QB 625.

91 See n 22 above.

92 See eg *Home Office v Dorset Yacht* [1970] AC 1004, per Lord Reid at 1030. The courts have tended to adopt an approach that involves a requirement that the harm must be both reasonably foreseeable and direct in cases where the harm was perceived to be reasonably foreseeable but indirect.

as will become apparent, the revised version of the risk theory permits the great majority of cases to be dealt with without recourse to that problematic concept.

Traditionally, the courts have invoked *novus actus interveniens* in cases, where, subsequent to the defendant's faulty conduct, there has been either deliberate human intervention, or some highly abnormal natural event, in the sequence of events leading to the claimant's injury. Now it should be apparent that, so far as natural events go, the revised risk theory straightaway renders reliance on the concept redundant. Instead, the finding that injury is too remote in such a case can be supported simply on the basis that the presence of the abnormal event as a causal condition in the set sufficient for the injury was not reasonably foreseeable.⁹³

Many of the cases that involve deliberate human intervention yield to the same analysis. That is to say, the risk that a third party might exploit the defendant's faulty act so as to inflict harm on the claimant was not one that, in the circumstances, a reasonable person would have guarded against.⁹⁴ Indeed, here the modified risk theory also serves to explain those converse instances where, notwithstanding deliberate human intervention, the courts *refuse* to hold that this amounted to a *novus actus*. Two cases in point are *Stansbie v Troman*⁹⁵ and *Reeves v Commissioner of Police for the Metropolis*.⁹⁶

In *Stansbie* a decorator failed to lock up the claimant's house on going out to buy some wallpaper. While the house was unoccupied, a burglar entered through the unlocked door and stole a diamond bracelet belonging to the claimant's wife. The Court of Appeal held the decorator liable in respect of this loss. More recently, in *Reeves* the House of Lords upheld a claim against the police in respect of their failure to prevent the suicide of a remand prisoner, known to be a suicide risk. This was despite the trial judge's finding that the man, at the relevant time, had been of sound mind.⁹⁷ It is clear that, in both cases, the possibility of deliberate human intervention (by the burglar and suicide, respectively) completing a causal set for harm was precisely the reason for regarding the defendant's conduct as faulty.⁹⁸

Thus far, the discussion has focused on the applicability of the doctrine of *novus actus interveniens* in the context of immediate harm suffered by the claimant. In such situations, the defendant's faulty conduct would have been insufficient (in the absence of the putative *novus actus*) for any harm at all. However, the doctrine has also been deployed so as to deny liability in certain ulterior harm cases. Thus, in *McKew v Holland and Hannen and Cubitts*,⁹⁹ the claimant, who suffered a broken ankle after his previously injured leg gave way, was not entitled to recover for this further injury against the defendant (who had caused the injured leg). The House of Lords held that the claimant's conduct in descending some stairs without assistance

93 Paradoxically, the revised risk theory can be seen as supportive of the broad intuition underlying the *novus actus* concept. For often what goes under that name will turn out to be a causal condition, whose presence in the causal set for harm was not reasonably foreseeable (eg the lightning strike in the Von Kries hypothetical, discussed at n 46 above).

94 Hart and Honore (n 3 above, at 74) mention the example of someone who carelessly starts a fire in the vicinity of the claimant's forest. The fire is about to go out, but an arsonist appears on the scene who pours petrol on it so that it spreads and destroys the forest.

95 [1948] 2 KB 48.

96 [2000] 1 AC 360.

97 The police, by leaving the wicket hatch to the prisoner's cell open (in express contradiction to suicide prevention guidelines), gave him the opportunity to attach a ligature, which he used to hang himself. Damages were reduced by 50 per cent, though, to reflect his contributory fault.

98 These are cases where, as noted above (see n 44), the hypothetical risk claim in respect of the defendant's conduct is predicated on the possibility (and necessity) of the further causal condition represented by the third party's voluntary act appearing in the causal set.

99 [1969] 3 All ER 1621.

(and knowing of his leg's propensity to give way), was unreasonable, and thus broke the chain of causation.

In fact, though, it is suggested that recourse to the notion of a *novus actus* in such a situation is unnecessary. As previously argued, cases where recovery for ulterior harm is allowed constitute an *exception* to the general rule that the claimant cannot recover for harm that was not reasonably foreseeable.¹⁰⁰ That exception operates in favour of claimants whose first injury compromised their resistance to the further injury in suit. Accordingly, in *McKew's* case, to hold the harm too remote it is enough to say that, at the time of the second accident, the claimant, who knowingly chose to take the risk, fell outside the exception.¹⁰¹

Are there, then, any circumstances in which the *novus actus* doctrine remains critical for the analysis of remoteness? It is submitted that one such situation may exist where, besides the defendant's faulty conduct, a *disproportionate criminal* act by a third party features in the causal set resulting in the claimant's injury. Admittedly, in a significant number of such cases the defendant will not have been under any duty of care to the claimant to protect him from such third party conduct.¹⁰² However, there are some exceptions to this general principle, where a duty will be imposed. These include situations in which the defendant had control over the third party and/or created a special danger of injury to the claimant.¹⁰³

Even so, to the extent that liability may be imposed on the defendant in such cases, it is suggested that it remains provisional and secondary in character. The third party is principally responsible for the harm, not only under the criminal justice system (in which, of course, the defendant's negligent provision of the opportunity for the crime is quite beside the point), but also in tort. This is evident if we consider the outcome if, say, a thief in a *Stansbie*-like scenario were caught and sued in civil law by the claimant for the loss of the bracelet: it is inconceivable that the thief could claim any contribution¹⁰⁴ from the person who left the door unlocked. Rather, in such cases, the secondary liability of the defendant arises only to cover the possibility of the criminal third party disappearing or turning out to be worthless.¹⁰⁵

It is against this background, that the special residual function of the *novus actus* doctrine should be understood. In particular, it represents the possibility of a defendant's plea that, though he may have been careless, his conduct 'paled into insignificance' besides that of the third party. As noted, the defendant is in reality a

100 See the discussion in n 81 above. In *McKew*, Lord Reid was of the view that the second accident suffered by the claimant was foreseeable. With respect, however, this is an example of the confusion between *theoretical* and *reasonable* foresight. The broken ankle may have been foreseeable in the former sense, but the risk of the claimant suffering the further injury would not have concerned a reasonable person who would not in any event have caused the initial injury, see n 77 above.

101 See also the facts of *Lamb v Camden LBC* [1981] QB 625, discussed at n 90 above.

102 Thus, though the defendant may reasonably appreciate that his conduct increases the risk of injury to the claimant, he cannot *a priori*, and as a matter of general legal policy, be held to account for such damage.

103 As in *Home Office v Dorset Yacht* [1970] AC 1004. The duty only extended to the special dangers presented to those nearby during the immediate escape of the boys, not to their general criminal acts against the public while at large. Another case arises where the defendant's fault has rendered the plaintiff vulnerable to the depredations of the third party: see the facts of *Stansbie v Troman*, n 95 above, and also the discussion at n 86.

104 ie, under the Civil Liability (Contribution) Act 1978.

105 His role is thus similar to that of an insurer of the harm, albeit he, unlike a normal insurer, has not entered the obligation voluntarily. Instead it is placed upon him because of his fault in providing the third party with the opportunity to commit the crime.

quasi-insurer for the third party, but while this may extend to cases where there is a close relationship between his fault and the nature of the subsequent crime, there will be other cases where the disproportion in the relative culpability of the two parties is so pronounced that we no longer feel it proper to hold the defendant to this role. Consider a variant of *Stansbie v Troman*, in which the decorator leaves the claimant asleep in the unlocked house, knowing that a dangerous criminal, who has a record of attacking people in their homes, is on the loose in the area. Here, even though the risk to the claimant, of being the victim of such an attack, may be regarded as reasonably foreseeable (though small, the risk is not far-fetched and the cost of the precaution is very low), it is submitted that a court would not wish to impose liability for it upon the decorator.

Conclusion

The argument in this paper has attempted to introduce a measure of clarity and consistency into an area of law that has long been regarded as obscure and *ad hoc*. The argument has proceeded in three main stages. First, the function of the enquiry into remoteness has been examined and related to the underlying issue of how it is that persons become responsible for outcomes in the world. Secondly, the concept of risk, central to the 'reasonable foreseeability' test, has been analysed to yield a revised version of the risk theory capable of resisting the accusation (by causal minimalists) that determinations of remoteness are entirely subjective and/or policy-driven. In this regard, it has been suggested that the 'reasonable foreseeability' test should be applied not to the 'outcome harm' in suit, but rather to the causal set that gave rise to the harm. Thirdly, while it has been conceded that in certain cases, and for reasons that are in a sense a matter of policy, legal responsibility may be extended (or curtailed), so that an agent's liability for an outcome is no longer coextensive with the risk that should reasonably have been foreseen, this does not undermine the general utility of the risk theory. Rather, such cases should be regarded as forming a limited group of exceptions to that theory.

Ultimately, and notwithstanding the confusion sometimes apparent on the surface of courts' decisions on remoteness of damage, it is submitted that this area of the law is far more principled than has commonly been supposed. Moreover, the relevant principles, springing as they do from deep-rooted notions of individual responsibility, accord with basic moral intuition.

Culture, Autonomy and *Djulibinyamurr*: Individual and Community in the Construction of Rights to Traditional Designs

Kimberlee Weatherall*

The possibility of granting proprietary rights in indigenous intangible cultural property – including artwork, cultural items and, more recently, traditional knowledge – has been and continues to be an area of considerable controversy, and the subject of discussions in various international organisations. It is widely accepted that present intellectual property regimes are structurally inadequate. The author focuses on the particular problem of traditional designs, and seeks to analyse critically the justifications that are advanced for extending existing regimes or introducing a sui generis right: in particular, the protection of 'cultural integrity'. The aim is to elucidate some of the theoretical problems with this rationale, and to extrapolate, from arguments regarding the importance of culture and cultural integrity, to the form and scope of rights that such an argument might require. In particular, the author believes that such a rationale has implications in determining how conflicts between communal and individual interests are to be resolved.

[T]here are situations which no amount of reflection will resolve and where thoughts act out, as it were, a conflict of interest. ... I could not bring my two minds together.¹

Thus did Marilyn Strathern describe the intellectual dilemmas regarding the application of intellectual property rights to traditional knowledge; her comments are equally applicable to the whole spectrum of indigenous intangible cultural property. When considering legal protection of indigenous 'folklore', now more commonly (and broadly) referred to as indigenous cultural and intellectual property,² one is invariably pulled in contradictory directions, both morally and logically. Strathern's statement captures well the inevitable sense of intellectual schizophrenia.

It is therefore hardly surprising that such issues have long been debated in numerous fora without resolution.³ Several factors have contributed to a recent increase in the international profile of this area: the rising value of indigenous art;⁴ publicity concerning (and protests over) efforts to exploit traditional knowledge by multinational pharmaceutical companies; World Intellectual Property Organization (WIPO) discussions regarding a possible international convention;⁵ the discussions

* LL.M Candidate, Yale University. The author would like to thank Dr Michael Spence of St Catherine's College, Oxford for his invaluable comments; any mistakes are the author's own. The author would also like to thank the British Foreign & Commonwealth Office and the Oxford-Australia Fund for their support at Oxford University

1 M. Strathern, 'Exploitable knowledge belongs to the creators of it' (1998) 6 *Social Anthropology* 109, 109.

2 A compendious term of contested content: M. Blakeney, 'What is Traditional Knowledge?' (1999) (WIPO/IPTK/RT/99/3) 2.

3 T. Janke, 'UNESCO-WIPO World Forum on the Protection of Folklore: Lessons for Protecting Indigenous Australian Cultural and Intellectual Property' (1997) 2 *Art, Antiquity & Law* 405, 406.

4 T. Janke, *Our Culture, Our Future* (1999) ch 2 (available at www.icip.lawnet.com.au).

5 WIPO Secretariat, *Protection of Traditional Knowledge, A Global Intellectual Property Issue*, WIPO/IPTK/RT/99/2 (22/10/1999) (note that this is an idea that WIPO refused to acknowledge as within its mandate only as recently as 1995)

concerning Article 8(j) of the Convention on Biological Diversity (CBD),⁶ and the heightened visibility of indigenous peoples' concerns worldwide, including assertions by non-government organisations (NGOs) and various indigenous fora of ownership of intangible cultural property.⁷

Many who advocate legal protection for indigenous cultural and intellectual property argue that a *sui generis* regime is required.⁸ There is broad agreement in the literature that Western intellectual property law does not fully protect such intangibles, particularly owing to structural features such as the duration of rights, requirements of originality (copyright) or novelty and inventive step (patent) and the focus on protecting the work of individual, identifiable authors or inventors.⁹ In addition, the whole structure of Western intellectual property law is arguably inconsistent with indigenous customary law, worldviews, and attitudes towards intangible property.¹⁰ Such structural issues have been extensively considered elsewhere. In this paper, I am concerned with analysing in some depth the *justifications* that are most commonly raised in support of the grant of proprietary rights in intangible cultural property, and the implications that the pursuit of such rationales have for the form and scope of such rights or interests. In particular, I will consider arguments founded on a concern with the protection of culture; the communal nature of the interests to which such a rationale leads. I then draw on the literature regarding group rights and group interests in looking at how such interests interact with other potential interests in the relevant cultural products.

Some provisos are required. First, in order to confine the issues, I will focus on traditional ('pre-existing'¹¹) designs in Australian Aboriginal art: designs handed down through generations of an Aboriginal community and re-embodied in new 'artworks' by individual artists. I thus distinguish between different subject-matters in a way many consider illegitimate;¹² nevertheless, I hope that at least some of this discussion will be relevant to other subject-matters.

6 Convention on Biological Diversity Executive Secretary, *Legal and Other Appropriate Forms of Protection for the Knowledge, Innovations and Practices of Indigenous and Local Communities*, UNEP/CBD/WG8J/1/2 (10/1/2000) (available at www.biodiv.org).

7 UN Draft Declaration on the Rights of Indigenous Peoples ('UNDDRIP'), Art. 29, 12; and indigenous declarations: *Mataatua Declaration on the Cultural and Intellectual Property Rights of Indigenous Peoples* (1993); *Kari-Oca Declaration* (1992); *Charter of the Indigenous-Tribal Peoples of the Tropical Forests* (1993); *Recommendations from the Voices of the Earth Congress* (1993); *Statement from the COICA/UNDP Regional Meeting on Intellectual Property Rights and Biodiversity* (1994); *Julayinbul Statement on Indigenous Intellectual Property Rights* (1993). These instruments are reprinted in D. Posey and G. Dutfield, *Beyond Intellectual Property: Towards Traditional Resource Rights for Indigenous Peoples and Local Communities* (Ottawa: International Development Research Centre, 1996).

8 Most submissions to Janke, n 4 above, favoured a *sui generis* response (§18.2); E. Daes, *Protection of the Heritage of Indigenous People* E/CN.4/Sub.2/1995/26 (United Nations Sub-Commission on the Prevention of Discrimination and Protection of Minorities, Final Report, 1995) Guidelines 13–15. *Sui generis* protection is not universally endorsed: K. Puri, 'Cultural Ownership and Intellectual Property Rights Post-Mabo: Putting Ideas into Action' (1995) 9 IPJ 293, 327. Arguments *against* include enforcement difficulties, and access to legal resources: M. Brown, 'Can Culture be Copyrighted?' (1998) 39 *Current Anthropology*, 193, 204; S. Brush, 'Whose Knowledge, Whose Genes, Whose Rights?' in S. Brush and D. Stabinsky, *Valuing Local Knowledge* (Washington DC: Island Press, 1995) 1, 10; R. Coombe, 'Intellectual property, human rights and sovereignty' (1998) 6 *Ind J Global Legal Stud* 59, 96.

9 J. McKeogh and A. Stewart, 'Intellectual Property and the Dreaming', in E. Johnson, M. Hinton and D. Rigney, *Indigenous Australians and the Law*, (1997) 53, 62; Puri, *ibid* 305 ff.

10 A. Paak, 'Cultural appropriation and the law: an analysis of the legal regimes concerning culture' (1993–4) 8 IPJ 57, 61–62.

11 D. Ellinson, 'Unauthorised Reproduction of Traditional Aboriginal Art' (1994) 17 UNSWLJ 327, 330.

12 E. Daes, *Protection of the Heritage of Indigenous Peoples* (UN Human Rights Series, UN Office of the High Commissioner for Human Rights, 1997) s 21; Declarations, n 7 above.

Second, I will not attempt further to define the protected subject matter. Serious problems arise here also: for example, in identifying and enforcing proprietary interests in designs lacking a material form; how to allow for incremental development without unduly restricting other artists (particularly non-indigenous artists); and finding an alternative to 'originality' to act as a threshold test. Further questions arise concerning the practical feasibility of protecting Aboriginal or other indigenous 'styles' (eg dot painting). These are exceedingly difficult issues; nevertheless, for present purposes, I will assume that the subject matter can be identified.

Third, the focus will be on issues in framing *sui generis* protection in the style of an intellectual property regime – providing rights in intangibles akin to property rights.¹³ Some commentators question whether private law remedies are applicable at all in this context;¹⁴ nevertheless, such a regime must be considered as one of the alternatives presently 'on the table'. There are other possibilities: in relation to traditional knowledge, for example, arrangements based on private contract and/or multilateral agreements have been mooted.¹⁵ Such approaches are *not* so widely advocated in the artistic field; perhaps in part because, at least *prima facie*, they are likely to be more useful where specific knowledge and (relatively) specific uses can be pre-identified, and where access to the relevant intangible brings users in direct contact with indigenous owners. A further option is a public trusteeship model.¹⁶

Finally, the Australian Aborigines whose designs are considered here are an *indigenous* people. Such peoples are characterised by continuity with pre-colonial populations, continuing distinctness from the rest of society, and determination to 'preserve, develop and transmit to future generations their ... ethnic identity, as the basis of their continued existence as peoples, in accordance with their own cultural patterns, social institutions and legal systems.'¹⁷ Given their distinct identity, and the special protection accorded indigenous peoples in international law, if a case can be made for recognition of communal interests in traditional designs, indigenous designs are the archetype. Extension of similar rights to tribal peoples, local communities, or minorities is a separate debate.¹⁸

The first section uses Australian materials to provide some background – exploring the nature of customary interests in traditional designs. The second section considers possible purposes served by protecting communal interests in such designs. The key reasons usually cited in this context are cultural – ie the need to protect certain elements of culture especially of indigenous groups. Conclusions from this section will then provide a framework for considering certain practical problems regarding the exercise of group property rights (the third section) and interactions between group and individual interests (the fourth section).

13 Cf Posey and Dutfield's 'traditional resource rights', n 7 above, which provide, not self-executing rights, but a framework of legal principles for appropriate relations with indigenous communities, and criteria for evaluating planned laws such as the *sui generis* regime considered here: n 6 above, § 50.

14 eg R. Coombe, 'Critical Cultural Legal Studies' (1998) Yale J L & Human 463.

15 Summarised by N. Roht-Ariza, 'Of seeds and shamans: the appropriation of the scientific and technical knowledge of indigenous and local communities' (1996) 17 Mich J Int'l L 919, 958 ff.

16 Blakeney, n 2 above, 10, 12. Public trusteeship-type models have been adopted in the copyright legislation of some African nations, see Kuruk, n 18 below.

17 M. Cobo, Special Rapporteur of the Sub-Commission on Prevention of Discrimination and Protection of Minorities, *Study of the Problem of Discrimination against Indigenous Populations* (E/CN.4/Sub.2/1986/7/Add.4) § 379; Blakeney, n 2 above, 3–4.

18 P. Kuruk, 'Protecting Folklore under Modern Intellectual Property Regimes: A Reappraisal of the Tensions between Individual and Communal Rights in Africa and the US' (1999) 48 Am ULR 769, 839.

Collective interests in traditional designs

The issue in this first section is this: how do Australian Aboriginal communities characterise interests in traditional designs, and what are their key concerns? There is no unified 'Aboriginal viewpoint' on these issues, nor do I purport to speak for Aboriginal people. The following¹⁹ draws particularly on evidence of indigenous community members in Australian copyright cases; and a recent report produced for the Aboriginal and Torres Strait Islander Commission (ATSIC).²⁰

Traditional designs in Australia

The cases

Several Australian cases have considered copyright issues relating to Aboriginal artworks incorporating traditional designs. The first of these proceedings, brought in 1989 by Aboriginal artists against a T-shirt manufacturer for copyright infringement in specific existing works of art, were settled in the artists' favour.²¹ In *Yumbulul v Reserve Bank of Australia*²² (a case concerning the design of the Australian ten dollar note), the relevant Aboriginal community from whom the artist derived his customary right to use particular designs asserted a communal interest in the artwork (a sculptural work in that case); the Court, however, could not find accommodation for such interests in Australian law, noting (perhaps in understatement) that 'it may ... be that Australia's copyright law does not provide adequate recognition of Aboriginal community claims to regulate the reproduction and use of works which are essentially communal in origin.'²³ Similar criticisms were expressed by the Court in *Milpurrruru v Indofurn Pty Ltd*,²⁴ a case concerning reproduction of artworks on carpets. *Milpurrruru* is significant chiefly because the Court awarded additional damages for cultural harm.²⁵

In *Bulun Bulun v R & T Textiles Pty Ltd*,²⁶ the Court was specifically asked to consider whether 'communal title in [Aboriginal peoples'] traditional ritual knowledge, and in particular their artwork, [could be] recognised and protected by the Australian legal system'.²⁷ The artwork, 'At the Waterhole', was of sacred significance. The waterhole itself, *Djulibinyamurr*, is the 'spring, life force and spiritual and totemic repository for [the artist's] lineage of the Ganalbingu people'.²⁸ The traditional designs depicted represented the 'number one item' of *madayin* (sacred or ritual knowledge).²⁹ The Court, asked to determine whether the artist *Bulun Bulun* held the copyright as a fiduciary and/or on trust for the Ganalbingu, found that the intention required to create an express trust was lacking.³⁰ However, a

19 The focus of this section is on traditional communities. I return to the issue of 'urban' indigenous artists below.

20 Janke, n 4 above.

21 C. Golvan, 'Aboriginal Art and Copyright: The Case for Johnny Bulun Bulun' [1989] 10 EIPR 346, 347.

22 (1991) 21 IPR 481 ('*Yumbulul*'). The actual results in this case turned on agency and contractual issues.

23 *ibid* 490.

24 (1994) 130 ALR 659 ('*Milpurrruru*').

25 *ibid* 692.

26 (1998) 157 ALR 193 ('*Bulun Bulun*').

27 *ibid* 195. That it was designed to be a test case for this particular issue is clear from the fact that the action was continued even after the complete capitulation of the respondent.

28 *ibid* 198.

29 *ibid* 199, 200.

30 *ibid* 206-207.

fiduciary relationship arose from the grant of permission by the Ganalbingu to Bulun Bulun to create the work, the use by the artist of ritual knowledge, and the requirement that that ritual knowledge be used in accordance with customary law.³¹ Thus customary law regarding communal interests was part of the factual matrix that allowed the Court to find a fiduciary relationship. The remedial consequences of this fiduciary relationship were confined, however, at least in that case to the internal relations between artist and community. Bulun Bulun as fiduciary was obliged not to exploit the artwork contrary to Ganalbingu law, and to take appropriate action to restrain and remedy copyright infringement.³² In the result, no relief was available since Bulun Bulun had taken appropriate action.³³

Themes: customary law and community

Several themes worth noting may be teased out of the cases and other literature. The first relates to the special, integral role traditional artwork fulfils in Aboriginal communal life and spiritual beliefs, and the deep attachment Aboriginal communities feel towards traditional designs. Traditional designs are a defining element of the communal and indigenous identity of Aboriginal groups;³⁴ an expression of the continuity of the community. Artworks are inseparable from the relationship between the community and its traditional land, so foundational to Aboriginal cultures (in common with many indigenous cultures).³⁵ In *Bulun Bulun*, evidence emphasised that the Ganalbingu were given designs by their Creator Ancestor *Barnda* together with the land, as 'part of [a] bundle of rights in the land [which] must be produced in accordance with Ganalbingu custom and law.'³⁶ Different designs serve different functions in communal life: for example ceremonial functions, the recording of history, culture and stories, and the education of younger generations.³⁷ For example, 'Morning Star Poles' like that featured in *Yumbulul* are used in ceremonies commemorating important persons' deaths, in gift exchange and inter-clan bonding. In Aboriginal traditional art, 'inside meaning' encoded in some artwork, recognisable only to the initiated, records communal ritual and law.³⁸ The whole community – not only particular designated 'artists' – is expected to reproduce traditional designs.³⁹ Thus traditional designs are integral to every aspect of the culture, communal life and identity; their embodiment in artwork in accordance with custom is considered essential to the maintenance of the community's vitality and traditions.⁴⁰

A second theme is the *communal* nature of interests in traditional designs, and the role of customary law in defining those interests.⁴¹ The complex kinship system central to Aboriginal customary law⁴² shapes rights and responsibilities regarding designs. The 'rights' in customary law in this context are more akin to

31 *ibid* 210.

32 *ibid* 211.

33 *ibid*.

34 Puri, n 8 above, 294–95, 297; Dees, n 12 above, s 21.

35 Puri, *ibid* 307; S. Gray, 'Wheeling, Dealing and Deconstruction: Aboriginal Art and the Land post-Mabo' (1993) 63 *Aboriginal Law Bulletin* 10, 11.

36 *Milparurru*, n 24 above, 663; *Bulun Bulun*, n 26 above, 198; *Yumbulul*, n 22 above, 483.

37 *ibid* 662.

38 *ibid* 662, n 26 above, 200.

39 Golvan, n 21 above, 348.

40 Puri designates traditional art 'social cement': n 8 above, 300.

41 I generalise; but it is important to realise that Aboriginal law is far from homogenous: Ellinson, n 11 above, 338; J. Toobey, 'Understanding Aboriginal Law' (unpublished; on file with author) 13.

42 Toobey, *ibid* 19.

'custodianship' than 'ownership',⁴³ a 'bundle of relationships, rather than a bundle of economic rights',⁴⁴ involving responsibility to past and future generations (in strong contrast to more usual Western notions of proprietorship).

Custom defines not only who may depict designs, but to whom designs, and their inner meanings may be revealed.⁴⁵ The degree and nature of customary controls, as well as who must be consulted before a design is used, depend on the design and the nature of the proposed use;⁴⁶ even widespread commercial use of some designs may be allowed after appropriate (though extensive) consultation.⁴⁷ An individual's entitlement under customary law to depict certain designs depends on a (varying) combination of factors such as age, descent, gender, initiation,⁴⁸ skill, and experience in the corpus of ritual knowledge.⁴⁹ Moreover, control is dispersed among different groups and individuals.⁵⁰ The right of any given artist to depict traditional designs is subject to various (non-exclusive) rights and responsibilities of other traditional owners,⁵¹ and must be exercised in compliance with custom and not contrary to the community's interests.⁵²

In short, the evidence reveals a complex matrix of interlocking rights and responsibilities.⁵³ In *Bulun Bulun* the artist, by right of seniority, initiation, skill, and kinship, was permitted to depict 'At the Waterhole'; another applicant, Milpururru, the artist's 'Djungayi' (a particular kind of kinship relation, of more than familial significance, involving as it does a certain level of responsibility⁵⁴), had the right to be involved in important decisions regarding its use; and consensus of all traditional owners was required for some uses.

The third theme concerns the role of the artist. Individual artists are not automatons controlled entirely by custom:⁵⁵ they have distinctive ways of expressing designs,⁵⁶ varying in skill, detail, and particular compositions of designs and techniques.⁵⁷ Thus although customary dictates are important,⁵⁸ development and incremental change are not thereby precluded.

Finally, two key concerns are expressed by Aboriginal people. The first is economic – there is a robust market for traditional and traditional-inspired artwork, and so unauthorised reproduction of works, or the construction of 'pastiche' works purporting to be 'Aboriginal' damages valuable economic interests. The second, more important set of frequently raised concerns relate to communal, cultural and spiritual integrity. Given the intimate connection between designs and communal identity emphasised above, it is not surprising that unauthorised reproduction causes deep offence;⁵⁹ such use is considered an attack on identity and the

43 Janke, n 4 above, §1.3.4; *Milpururru*, n 24 above, 663.

44 *Dacs*, n 12 above, §26.

45 *Milpururru*, n 24 above, 662, 664.

46 *ibid* 664.

47 *ibid*.

48 *Yumbulul*, n 22 above, 483.

49 *Milpururru*, n 24 above, 662.

50 *ibid* 663; Ellinson, n 11 above, 335–337.

51 J. Weiner, 'Protection of Folklore: A Political and Legal Challenge' (1987) 18 IIC 56, 72; *Milpururru*, *ibid* 663.

52 Janke, n 4 above, §1.3.4; *Dacs*, n 12 above, §29; *Bulun Bulun*, n 26 above, 209.

53 K. Maddock, 'Copyright and traditional designs – an Aboriginal dilemma' (1989) 2 IP 7, 8–9.

54 Explained by Ellinson, n 11 above, 335–336; *Bulun Bulun*, n 26 above, 199.

55 Janke, n 4 above, §1.3.3, §1.8.4.7; Department of Home Affairs and Environment, *Report of the Working Party on the Protection of Aboriginal Folklore* (Canberra: AGPS, 1981) ('Folklore Report') §505.

56 *Golvan*, n 21 above, 349–350.

57 *Milpururru*, n 24 above, 665.

58 *ibid* 662.

59 *ibid* 663; *Yumbulul*, n 22 above, 483; *Bulun Bulun*, n 26 above, 199.

community, an interference in the relationship between the artist, their ancestors and the Creator Ancestor(s).⁶⁰ *Bulun Bulun* summarised these concerns:

Unauthorised reproduction of 'At the Waterhole' threatens the whole system and ways that underpin the stability and continuance of Yolngu society. It interferes with the relationship between people, their creator ancestors and the land given to the people by their creator ancestor. It interferes with our custom and ritual, and threaten[s] our rights as traditional Aboriginal owners of the land and impedes in the carrying out of the obligations that go with this ownership and which require us to tell and remember the story of Barnda, as it has been passed down and respected over countless generations.⁶¹

Communal and customary control is considered necessary to protect the community and its vitality.

Customary laws vary widely between indigenous peoples and between countries,⁶² and not all interests in indigenous cultural items assume a collective form.⁶³ Nevertheless, strikingly similar themes and concerns are expressed in the international sphere. UN Special Rapporteur Erica-Irene Daes found that indigenous people view heritage not as property but rather 'in terms of community and individual responsibility'; and as 'a communal right, associated with a family, clan, tribe or other kinship group'.⁶⁴ Indigenous declarations,⁶⁵ recent WIPO studies⁶⁶ and other writings⁶⁷ highlight not only economic concerns, but more importantly the integral role that indigenous cultural and intellectual property fulfils in indigenous communities, and the importance of control over cultural items in maintaining indigenous identity, communal life and vitality.⁶⁸

Return to Bulun Bulun

I will not consider in detail the application of trust or fiduciary law in *Bulun Bulun*.⁶⁹ *Bulun Bulun* was an attempt to gain explicit recognition of customary communal interests in traditional artwork by drawing analogies with Western concepts – and, no doubt, to highlight inadequacies of the law in this area. Analogies may be helpful, but owing to differences regarding concepts of property and ownership, we must question the appropriateness of imposing cultural values embedded in our legal system on the resolution of such claims.⁷⁰ Certainly classical Western intellectual property law or more general property concepts do not capture the *cultural* importance of artwork and designs. Both trust and fiduciary law envisage a concentration of rights, duties and responsibilities in the fiduciary/trustee;⁷¹ and are not designed to regulate the complex of interlocking, non-exclusive, contingent rights and responsibilities described above. Most

60 *Bulun Bulun*, *ibid* 199.

61 *ibid*.

62 Four Directions Council, quoted in G. Dutfield, 'Between a rock and a hard place: indigenous peoples, nation states and the multinationals' (available www.fao.org/docrep/w7261e/w7261e06.htm).

63 Coombe, n 8 above, 108; S. Harding, 'Value, Obligation and Cultural Heritage' (1999) 31 *Ariz St L J* 291, 305–306.

64 Daes, n 12 above, ss 26, 28.

65 n 7 above.

66 eg WIPO Secretariat, *Report of the WIPO Fact-Finding Mission on Traditional Knowledge, Innovations and Culture to Central America* (WIPO/FFM.C-AM/IMR/99/5) (20/10/1999), 5 (www.wipo.org).

67 Daes, n 12 above, ss 30, 115, Principle 171; R. Coombe, 'The Properties of Culture and the Politics of Possessing Identity' (1993) 6 *Can J L & Juris* 249, 284.

68 *ibid*, s 30; Kuruk, n 18 above, 781–786.

69 See A. Kenyon, 'The 'Artist Fiduciary' – Australian Aboriginal Art and Copyright' [1999] *Ent L R* 42.

70 C. Bell, 'Aboriginal Claims to Cultural Property in Canada' (1992) 17 *Am Ind LR* 457, 462.

71 cf the Ghanaian tribal head, an example cited by the Court, who resembles a trustee more than an Aboriginal artist: *Bulun Bulun*, n 26 above, 209.

importantly, even equitable ownership of copyright by a community cannot provide interests in the traditional designs themselves.

No doubt some aims of the litigation were achieved; nevertheless, equitable ownership of copyright is useful chiefly as a 'stop-gap' measure,⁷² and to force the Court to recognise the *existence* (if not the full force) of customary communal interests.⁷³ The characterisation of the artist as fiduciary has these limited practical benefits, but on another level is also deeply problematic – potentially, it allows Courts applying fiduciary law to intervene in intra-community relations and even potentially to enforce (their interpretation of) customary law, but at the same time offers no protection against the actions of non-members of the community not bound by customary obligations.

Justifications for legal protection of traditional designs

Structural reasons for copyright law's inadequacy in protecting traditional designs are examined in more detail elsewhere.⁷⁴ In considering whether (and how) a *sui generis* regime may address the communal interests and concerns outlined in the first section of this article, we must consider what justifications, arising from such communal interests and concerns, could or would support such a regime of communal rights to (a) control use⁷⁵ and (b) derive profit from the exploitation of traditional designs.

Traditional designs presently fall within the 'public domain'. As a result, even those who are sympathetic towards the cause of indigenous peoples thus run up against the not insignificant problem that allowing any such claim leads to further 'partitioning' of the cultural 'commons'. A claim to exclusive access to part of that 'pool' of otherwise unprotected ideas and expressions restricts other people's liberties to express themselves in certain ways, and hence requires strong justification.⁷⁶ 'Mere attachment' to designs, even if deeply felt – ie, the fact that designs are significant, even sacred to certain people – cannot be enough: if all such designs were protected, the public domain would be obliterated.

Cultural integrity

We have many particular things which we hold internal to our cultures. These things are spiritual in nature ... They are *ours* and they are *not* for sale. ... such matters are our 'secrets', the things which bind us together in our identities as distinct peoples. It's not that we never make outsiders aware of our secrets, but *we* – not *they* – decide what, how much, and to what purpose this knowledge is to be put. That's absolutely essential to our cultural integrity, and thus to our survival as peoples.⁷⁷

The more important concerns of indigenous people expressed in the first section were 'cultural' concerns – the desire to nurture cultural vitality, and to protect culture from interference. These concerns may be encapsulated in the concept of

72 cf C. Golvan, 'Aboriginal Art and the Protection of Indigenous Cultural Rights' [1992] 7 EIPR 227, 230. Golvan considers the possibilities of equitable ownership.

73 S. Gray, 'Black enough? Urban and non-traditional Aboriginal art and proposed legislative protection for Aboriginal art' (1996) 7 *Culture and Policy* 29, 39.

74 Puri, n 8 above; Ellinson, n 11 above; Janke, n 4 above, ch 5.

75 cf Folklore Report, specifically rejecting such a right: n 55 above, s 1221(b).

76 J. Waldron, 'From authors to copiers: individual rights and social values in intellectual property law' (1993) 68 *Chi-Kent LR* 841, 887.

77 Quoted in Coombe, n 67 above, 279.

'cultural integrity'. It is argued that control over the use of traditional designs is essential to maintain the cultural and spiritual integrity of an indigenous community and that this is sufficient reason to grant rights of control over traditional designs to the community.

What is 'cultural integrity', and why is it valuable?

'Cultural integrity' is a contested and problematic concept. 'Culture' as used here refers to the culture of a community as a whole, including traditional ways of life, communal institutions, languages, and distinct forms of cultural expression.⁷⁸ This broad definition reflects the concerns and understandings expressed by indigenous advocates themselves. 'Integrity' denotes the continued distinct existence of that culture without unwanted interference. The term 'cultural integrity' is used by advocates, and in international arenas, in which indigenous peoples are increasingly asserting, in 'rights-language' and in the language of property, their *rights* to cultural integrity and to control their cultural heritage,⁷⁹ frequently as one aspect of wider claims to self-determination.⁸⁰ Simple assertion however of such broad, categorical 'rights' or moral values does not advance our understanding very far: assessing the practical implications, and their weight in the inevitable conflicts with other interests, requires us to consider more critically why cultural integrity is *needed*, and the purposes it may serve. As the present concern is rights and interests in intellectual and cultural property, I will not enter into the broader controversy regarding the meaning of 'self-determination'. Whether or not 'cultural integrity' is an *aspect* of self-determination,⁸¹ it warrants separate consideration.

The benefits of cultural integrity, and cultural diversity more generally, as well as the special claims of indigenous and traditional peoples have been averred in international debates.⁸² Addressing only one aspect traditional designs the Court in *Bulun Bulun* asserted:

The evidence is all one way. The ritual knowledge ... embodied within the artistic work is of great importance to members of the Ganalbingu people. I have no hesitation in holding that the interest of Ganalbingu people in the protection of that ritual knowledge from exploitation which is contrary to their law and custom is *deserving* of the protection of the Australian legal system.⁸³

Such arguments are intuitively powerful – undeniably, equal respect is due to indigenous cultures⁸⁴ – but what is the source of this 'desert'? Benefits to society from simple 'cultural diversity' would not justify special protection for any

78 One of Coombe's three meanings of 'culture': n 8 above, 73–5; D. Réaume, 'Justice between cultures' (1995) 29 UBC L Rev 117, 120. Even the definition of 'culture' is debated: Coombe, *ibid*.

79 Exact terminology varies: J. Anaya, 'The Capacity of International Law to Advance Ethnic or Nationality Rights Claims' in W. Kymlicka (ed), *The Rights of Minority Cultures* (Oxford: OUP, 1995), 325; F. Naqvi, 'People's Rights or Victim's Rights: Re-examining the Conceptualization of Indigenous Rights in International Law' (1996) Ind L J 673, 714 ff; Declarations, n 7 above. Some use 'cultural integrity' referring only to the integrity of cultural/artistic productions: eg Janke, n 4 above, s 3.15; *Bulun Bulun*, n 26 above, 210. Such a reading is more narrow than the concerns considered here.

80 eg *Mataatua Declaration* (Preamble); *COICA Statement* Art 3, n 7 above.

81 The broad definition of 'culture' brings 'cultural integrity' closer to questions of self-determination: both are concerned to some extent with independence and continuation of autonomous institutions; however, debates regarding 'self-determination' tend to focus on *political* self-determination, rather than the broader cultural concerns considered here.

82 *Daca*, n 8 above, Principle 1; UNDDRP, n 7 above, Art 29.

83 *Bulun Bulun*, n 26 above, 210–211 (emphasis added).

84 Art 1, *UNESCO Declaration on the Principles of International Cultural Cooperation* (1966).

particular culture.⁸⁵ Arguments such as these, applied universally, would lead to a highly balkanised society and, as suggested earlier, the complete privatisation of the cultural commons.

A number of theories purport to explain the values enhanced by the continued existence of distinct cultures. One strong argument is based on the importance of group membership to individuals: culture, and in particular one's *own* culture, is essential to self-identification,⁸⁶ and to individual autonomy; it provides the context and alternatives essential to facilitate the pursuit of individual conceptions of the good.⁸⁷ Réaume on the other hand argues that an individual-centric approach fails adequately to address the collective (non-individualisable) nature of the interest in culture.⁸⁸ Garett's more radical approach asserts the intrinsic value of 'groupness'⁸⁹ as an essential component of human life.

The nature of the interests involved, and in particular, determination of the question whether the value of cultures is founded on individual benefit, or on the intrinsic good of the group, may determine priorities between group and individual rights.⁹⁰ Approaches based on individual benefit tend to prioritise individual rights over group interests; 'group good' approaches allow, at least in some circumstances, priority to the collective interest. Such questions are hotly contested, and the final choice between approaches may depend on one's politico-philosophical approach. It is not possible to resolve this theoretical debate here, but it is important to recognise its existence: these issues will resurface later when I come to consider conflicts between group and individual interests. In terms of providing a foundational justification for a *sui generis* regime, most important is the 'brute fact'⁹¹ that people's sense of identity, and self-respect, *are* bound up with their group cultures. This is consonant with the concerns referred to at p. 221 above. Equally clear is that the interests concerned have a collective aspect; they cannot be enjoyed alone; their value lies in their collective creation and enjoyment.⁹²

Such theories would not justify protective measures for every culture, even if effectively subsumed within a dominant culture. Rather, active legal protection which affects non-members' interests is most readily justified in the case of 'societal' cultures:⁹³ those which provide members with 'meaningful ways of life across the full range of human activities, including social, educational, religious, recreational, and economic life, encompassing both public and private spheres'⁹⁴ – those with the most comprehensive influence on individuals alone and collectively, and so those most likely to be missed if removed. In these cases, protection may be justified even at some social cost. Indigenous communities, who are characterised by their separateness from the dominant society and their determination to remain distinct, are (at least in some cases) the archetype of a societal culture.

85 Réaume, n 78 above, 119

86 A. Margalit and J. Raz, 'National Self-Determination' in Kymlicka, n 79 above, 79, 87; J. Raz, 'Multiculturalism: A Liberal Perspective' (1994) *Dissent* 67, 71.

87 W. Kymlicka, *Multicultural Citizenship: A Liberal Theory of Minority Rights* (Oxford: OUP, 1995), 82–93.

88 n 78 above, 130; L. McDonald 'Can collective and individual rights co-exist?' (1998) 22 *Melb ULR* 310.

89 R. Garett, 'Communitarianism and Existence: The Rights of Groups' (1983) 56 *S Cal LR* 1001; T. Dougherty, 'Group Rights To Cultural Survival' (1998) 29 *Columbia Human Rights Law Review* 355.

90 D. Johnston, 'Native Rights as Collective Rights: A Question of Group Self-Preservation' in Kymlicka, n 79 above, 179, 188ff.

91 Margalit and Raz, n 86 above, 87.

92 McDonald, n 88 above, 317–318.

93 Kymlicka's term. cf Johnston, n 90 above, ('multi-dimensional'); Margalit and Raz, n 86 above ('pervasive cultures').

94 Kymlicka, n 87 above, 76.

It is important to note that the reasoning considered here supports the protection of a distinct but presently existing culture – not the preservation of particular historical practices. Protection of cultural integrity means protection, not of the historical ‘authenticity’ of particular practices, but of the distinct existence of a multi-dimensional culture. It is their cultures as whole, living cultures, that indigenous peoples assert a right to defend.

Relationship with traditional designs

A broad concern with cultural integrity may justify a law granting exclusive rights in traditional designs to an indigenous community, depending on their significance. If designs occupy a central role in a societal culture, and if, as is argued,⁹⁵ their capacity to fulfil that role is damaged by use of designs contrary to customary law, or which dilutes the meaning and significance of important designs, or which deeply offends the community, eg through disrespect to sacred designs, and if the resulting cultural damage threatens or seriously harms the ongoing distinct existence of the culture (immediately or cumulatively), then *sui generis* legal protection is arguably justifiable to protect cultural integrity. This admittedly high threshold of seriousness is appropriate since even abuse of traditional designs does not directly raise serious questions of oppression, or forced assimilation.⁹⁶ In a mixed society, not every affront to cultural integrity or identity will warrant legal restriction.

It is almost impossible empirically to *prove* cultural harm, which is likely to be incremental.⁹⁷ There is however abundant supportive evidence, some of which has been described above: eg the offence caused by inappropriate use, the distress which has caused some Aboriginal artists to temporarily cease painting,⁹⁸ and views expressed by indigenous people as to the importance of cultural integrity.⁹⁹ Whether that evidence is sufficient to suggest that cultural integrity is threatened is a matter for assessment, and will depend on surrounding circumstances. That it *could* have such an effect seems clear.

A further argument which links cultural integrity to protection for traditional designs is that such protection is required in order to respect customary law, by recognising the ways in which customary law allocates control over designs.¹⁰⁰ The courts in the Australian cases frequently comment on the failure of Australian law to recognise customary interests in designs.¹⁰¹ Such statements recognise the legitimate existence and ongoing influence of customary law for community members.¹⁰² Respect for, or recognition of, an important institution central to the distinctness of a societal culture supports the right of members to its continued existence, and will undoubtedly enhance its ongoing relevance. This argument is the strongest in favour of a *sui generis* system granting rights of control, since it is at least arguable that the right to be protected from offence would not support a full property right.

⁹⁵ At 219–221 above; A. Jabbour, ‘Folklore Protection and National Patrimony’ (1983) 18 *Copyright Bulletin* 10, 11; Puri, n 8 above, 295.

⁹⁶ The threshold may vary with the level of legal protection proposed; relatively stringent property rights considered here justify a high threshold.

⁹⁷ Janke proposes a study: n 4 above, recommendation 2.2.

⁹⁸ Golvan, n 72 above, 228.

⁹⁹ Janke, n 4 above, s 3.15.

¹⁰⁰ Ellinson, n 11 above, 343; Puri, n 8 above, 316 (arguing traditional designs are property, relying on customary law); *COICA Statement*, *Julayinbul Statement*, n 7 above. McDonald offers a similar argument in relation to native title: n 88 above, 317.

¹⁰¹ eg *Yumbulul*, n 22 above, 490, *Milpurrurru*, n 24 above, 664–665; *Bulun Bulun*, n 26 above, 195.

¹⁰² Consonant with native title decisions: *Mabo v Queensland (No 2)* (1992) 175 CLR 1 (*‘Mabo’*).

Strengths and weaknesses of the cultural integrity argument

The cultural integrity rationale has several merits. First, it directly addresses concerns with cultural integrity and communal identity expressed by indigenous peoples. Second, it specifically justifies *communal* rights in traditional designs. If the aim is to support the community's culture, then the community's own (customary) rules must determine use of traditional designs. That interests in cultural integrity must be communal is also clear from the fact that such interests are 'non-individualisable'.¹⁰³ The individual of course has an interest in cultural integrity but not one that can be enjoyed alone.¹⁰⁴ Third, cultural integrity provides guiding principles for framing legal protection: any departure from customary rules tends to undermine the purpose of protecting cultural integrity, and requires independent justification. Finally, it provides some answer to a concern raised in the literature,¹⁰⁵ namely, how to limit the groups that can claim designs – although problems distinguishing those cultures which are 'societal' will remain.

The cultural integrity rationale supports a right to *control* use of traditional designs, including the right to prevent inappropriate commercialisation. It less obviously warrants a positive right to extract economic benefits. Practically speaking, rights to control lead inevitably to rights of exploitation: the alternative would be complete withdrawal of traditional designs from the market, with consequent detriment to society as a whole. A further indirect argument is that cultures survive through use:¹⁰⁶ for traditional designs, this means keeping alive meanings, traditional techniques, and customary laws. Economic incentives may support independence and encourage intergenerational transfer of skills,¹⁰⁷ thus furthering the ultimate aim of supporting distinct cultures.¹⁰⁸ Individual incentives are likely only partially to achieve such purposes,¹⁰⁹ and risk giving priority to personal gain over community interests. One study even suggested that, in remote Aboriginal communities, significant art production is maintained only where there is access to markets.¹¹⁰

One criticism of the right to extract economic benefit and Western intellectual property laws in general is that commercialisation corrupts, or irrevocably changes a culture, rather than protecting its 'distinctness'.¹¹¹ While the sentiment thus expressed sounds paternalistic, the effect that full property rights have on a culture is a relevant consideration. On the other hand, some method for dealing with the reality of the capitalist economy is required: rights of control may even be necessary to *prevent* commercialisation.¹¹² Property can forestall a free-for-all, allowing the group to assert control as against the world, while maintaining distinct intra-group relations of responsibility and obligation. In fact, various declarations

103 Réaume, n 78 above, 124–5.

104 J. Waldron, 'Communal goods as human rights' in J. Waldron, *Liberal Rights: Collected Papers* (Cambridge: Cambridge UP, 1993) 339, 354–358.

105 eg M. Brown, 'Reply' (1998) 39 *Current Anthropology* 218, 218.

106 Coombe, n 8 above, 93.

107 See similar arguments in the CBD: UNEP/CBD/TKBD/1/2 ss 11, 62.

108 C. Jacoby and C. Weiss, 'Recognizing Property Rights in Traditional Biocultural Contribution' (1997) 16 *Stan Envtl LJ* 74, 94.

109 Coombe, n 8 above, 109.

110 Department of Aboriginal Affairs, *The Aboriginal Arts and Crafts Industry* (Canberra: AGPS, 1989) 286–287. This finding cautions against ignoring interactions of different influences, market and non-market, although it can be explained consistently with sources emphasising cultural reasons for art production (eg broader appreciation contributes to art production by stimulating pride in culture; economic independence facilitates cultural production).

111 S. Gudeman, 'Sketches, Qualms, and Other Thoughts on Intellectual Property Rights' in Brush and Stabinsky, n 8 above, 102, 112.

112 M. Da Cunha, in Strathern, n 1 above, 112.

of indigenous peoples display a consciousness of the danger of 'commodifying' culture;¹¹³ in Australia, custom has often led to 'appropriate' uses eg for display in museums or as 'high art'.

There are however two more serious problems with 'cultural integrity' as a rationale for the grant of proprietary rights.

The first raises difficult and controversial political issues. The point of cultural integrity is that it justifies legal protection which ensures that a societal culture can remain distinct from the broader society. Is this not 'apartheid of the mind'?¹¹⁴ Similar arguments in other contexts have justified apartheid-like special provisions such as separate schools or administrative systems. Minow warns of the dangers of 'crude notions of identity': 'balkanisation, fragmentation, fundamentalism, illiberalism, segregation and prejudice'¹¹⁵ (not to mention racial discrimination). However, control over traditional designs is unlikely in itself to fragment society: it may be the bare minimum that can be offered, positive in effect and limited in impact on non-members. On the other hand, the association drawn by indigenous people between control of intangible cultural property and questions of self-determination heighten the potentially divisive effect.¹¹⁶ Such issues, of course, are not exclusive to traditional designs; there is no space here to enter into the broader political debate. The divisive potential of cultural integrity is a danger to be taken into account. The wise law-maker in this politically charged arena walks a tight-rope between proper protection for minority interests, and accusations of apartheid and paternalism.

The second challenge however goes to the validity of the underlying core concepts. The concept of 'culture' defies the kind of definition that would support rigorous analysis. The essence of the second challenge is that cultures have no boundaries or fixed existence – they influence and are influenced by other cultures, and shade into one another – so that integrity is meaningless and culture itself too fluid and contested to be a valid basis for legal distinctions. On this view, any attempt to found legal rules on cultural integrity reifies a purified, static image of culture.¹¹⁷

To put it crudely, we need culture, but we do not need cultural integrity ... In general, there is something artificial about a commitment to *preserve* minority cultures. Cultures live and grow, change and sometimes wither away; they amalgamate with other cultures, or they adapt themselves to geographical or demographic necessity. To *preserve* a culture is often to take a favored 'snapshot' version of it, and insist that this version must persist at all cost, in its defined purity, irrespective of the surrounding social, economic, and political circumstances.¹¹⁸

Waldron argues that any attempt to protect cultures insulates them from forces that allow them to operate in a context of genuine choice¹¹⁹ – contradicting the

113 D.A. Posey, 'Comment' (1998) 39 *Current Anthropology* 211.

114 Brown, n 8 above, 204.

115 M. Minow, 'The Constitution and the Subgroup Question' (1995) 71 *Ind LJ* 1, 24; P. Descola, 'Comment' (1998) 39 *Current Anthropology* 208; T. W. Pogge, 'Group Rights and Ethnicity' in I. Shapiro and W. Kymlicka, *Ethnicity and Group Rights* (39 *Nomos*) (New York: New York University Press, 1997), 188–189.

116 Declarations, n 7 above; T. Simpson, *Indigenous Heritage and Self Determination: the cultural and intellectual property rights of indigenous peoples* (IWGIA, 1997) s 8.9; Blakeney, n 2 above, 14–15.

117 Coombe, n 67 above, 266; S. Gray, 'Squatting in Red Dust: Non-Aboriginal Law's Construction of the "Traditional" Aboriginal Artist' (1996) 14 *Law in Context* 29, 32.

118 J. Waldron, 'Minority cultures and the cosmopolitan alternative' in Kymlicka, n 79 above, 101, 109–110.

119 *Ibid* 110.

underlying rationale in so far as it facilitates autonomy.¹²⁰ Furthermore, protection of a culture's 'essence'¹²¹ has paternalistic overtones.

Indigenous groups have responded by criticising the focus on authenticity, and emphasising their right to protect their distinct, not simply authentic cultures.¹²² Characterisation of cultures as societal in the sense that they have a complex, present influence presumes living, changing cultures; the underlying aim of cultural integrity need not refer to the traditional authenticity of particular aspects of the culture, like artwork, but rather the protection of the culture's distinctness.¹²³

But such responses do not fully answer the key criticism: that culture is too amorphous to found enforceable legal rights. This issue divides the literature. To deny the distinctness particularly of indigenous cultures, or their importance to members, is to deny reality – and claims that are fundamental to the people concerned. To the extent that they can be identified, distinct cultures, with their strong potential to affect individuals' lives, are worth protecting,¹²⁴ if it is possible without impermissibly encroaching on other interests. But it is equally unconvincing to assert that distinct cultures can be unambiguously defined. Perhaps the only conclusion – however unsatisfying – is that the concept of a distinct societal culture is one end of a spectrum where the other pole is Waldron's cosmopolitanism. Any law actually framed will likely fall back on approximations; particular care would be required to ensure that such concepts did not unduly stray from the underlying justifications. The wise law-maker's tightrope is not only thin but obscure and constantly shifting.

Other justifications

A number of other arguments focusing on economic interests are found in the literature. On examination, it is clear that these arguments are insufficient on their own to support full property-style rights of both control and economic exploitation.

Free-riding

One common argument may be loosely characterised as 'unjust enrichment': it is unjust, or unfair that commercial users profit using traditional designs without benefits going to traditional owners.¹²⁵ Such arguments have obvious instinctive appeal, however, some underlying factor is required to make such commercial profiteering 'wrong' – since we all frequently reap where we have not sown,¹²⁶ particularly in using public domain material. We need to understand *why* the community is the better claimant.¹²⁷ The first reason is that the community has the better claim to designs that are significant to its culture, or because it has such interests in customary law – this clearly relies on the same cultural integrity arguments outlined above.¹²⁸ The second possible basis is that the community has

120 At 224 above.

121 M. Minow, 'Identities' (1991) 3 *Yale Journal of Law and the Humanities* 97, 97–8; L. Prott, 'Cultural Rights as People's Rights in International Law' in J. Crawford, *The Rights of Peoples* (Oxford: OUP, 1988) 93, 95.

122 Janke, n 4 above, s 18.4.7; Réaume, n 78 above, 133–135.

123 Naqvi, n 79 above, 726.

124 Different considerations might arise in relation to oppressive cultures.

125 Jabbour, n 95 above, 12; Puri, n 8 above, 296; Weimer, n 51 above, 80.

126 W. Gordon, 'On Owning Information: Intellectual Property and the Restitutory Impulse' (1992) 78 *Va LR* 149, 178–180.

127 M. Spence, 'Passing Off and the Misappropriation of Valuable Intangibles' (1996) 112 *LQR* 472, 489 ff.

128 Such arguments are insufficient to sustain property rights, but may supplement a 'cultural integrity' argument, explaining why communities are entitled to full market value.

the better right to the designs as products of the community in the intergenerational sense. The difficulty here is that the classic Lockean argument 'I made it therefore it's mine'¹²⁹ is less compelling as an argument based on merit or desert when reframed 'our ancestors made it, therefore it's ours'. The element of property through labour that supports a classical Lockean argument is absent. Nor could a desert-based argument justify intergenerational property claims. Moreover, even if we accept that the community in some broader sense did 'create value' in traditional designs, that does not lead ineluctably to property rights, and particularly control.¹³⁰ Other forms of compensation or benefit-sharing are also possible.

Economic harm

An alternative argument that aims at the same alleged 'wrong' is the claim that commercial use by non-members of the community causes economic harm to the Aboriginal community – for example, through mass-produced cheap imitations,¹³¹ or by crowding the market and damaging opportunities (and prices) for traditional artists. Again, an antecedent basis for a community claim is required.

Broader distributive concerns

Another argument is that existing intellectual property regimes protect Western intellectual products, but systematically ignore or undervalue the intangibles of indigenous peoples¹³² or (to extend the argument further) protect the products of developed, not developing nations.¹³³ On this view, reciprocity requires the same recognition be given to the intellectual products of different cultures.¹³⁴ Again, despite the clear moral force of the argument, there are difficulties in using this argument to justify property rights in this case.

First, the application of 'special measures'¹³⁵ is (unfortunately) an easy target for critics who will claim that it extends the rights of indigenous people beyond those accorded to other individuals.¹³⁶ This is an argument for 'formal equality', and subject to the usual criticisms of formal equality (which need not be rehearsed here), though experience suggests it may have considerable political appeal. A stronger counter-argument is that although items such as traditional designs have some features in common with subject-matters protected by intellectual property law, there are also crucial differences, such as the lack of material form, which may justify different legal treatment and found an argument that, while compensation or 'equitable benefit-sharing' is required, property rights are not, owing to the indefinite nature of that which is claimed.¹³⁷ Furthermore, the rights claimed by indigenous people are different in scope from Western intellectual property law:

129 E. Hettinger, 'Justifying Intellectual Property' (1989) 19 *Phil & Pub Aff* 31, 38–39.

130 Gordon, n 126 above 167.

131 Janke, n 4 above, s 3.15.4.

132 A. Gupta, 'Building on what the poor are rich in' (www.csf.colorado.edu/aristi/papers/building.html); Kuruk, n 18 above, 775.

133 Gupta, *ibid*; E. Marden, 'The Neem Tree Patent: International Conflict over the Commodification of Life' (1999) 22 *Boston College International and Comparative Law Review* 279, 295.

134 cf C. Rose: *partial* proprietisation leads to conflict and injustice: 'The Several Futures of Property: of Cyberspace and Folk Tales, Emission Trades and Ecosystems' (1998) 83 *Minn LR* 129, 152.

135 UNDDRIP, n 7 above, Art. 29.

136 This argument was raised by the US against the present form of Art. 29: *An Analysis of the UN Draft Declaration on the Rights of Indigenous Peoples* (1999) (available www.atix.gov.au).

137 CBD signatories were able to agree to 'equitable benefit sharing' with indigenous peoples, but have not agreed to appropriate mechanisms: Convention on Biological Diversity, *Report of the Panel on Access and Benefit Sharing* (UNEP/CBD/COP/5/8). I choose the lack of material form because this feature is one which renders it very difficult to identify the precise subject matter of any proposed property rights.

for example, indigenous peoples claim rights of perpetual duration.¹³⁸ An argument from reciprocity or equal treatment would not support such a claim.

'Restorative'¹³⁹ arguments

A final rationale is that as a result of colonisation, indigenous peoples have suffered the loss of most resources including land; they therefore have a stronger claim to cultural expressions, as one of their few remaining resources.¹⁴⁰ This is a policy argument of great moral force, but does not support any particular form of rights. There are many ways in which historical inequities and disadvantage can and should be addressed.¹⁴¹

I wish to make it clear that I am not dismissing the arguments criticised above as irrelevant. The point is that despite the moral force of such arguments, the reasoning tends ineluctably to return to the underlying rationale of protecting culture, as it is only this aspect which distinguishes the claims of indigenous peoples from any other groups who may be economically harmed by some public domain activity and 'appropriation'. There is truth in the cliché that property does not arise merely from the existence of (monetary) value. The above considerations are relevant to political debate as strong additional policy arguments lending moral weight to the claim for protection of traditional designs.¹⁴² Alternative justifications based on economic considerations are inadequate to support the rights of control at the heart of indigenous demands. Stronger reasons founded in culture are required, and since mere attachment cannot be sufficient, we fall back on arguments concerning the protection of cultural integrity – which are strongest where 'societal' cultures are involved. The corollary is that so long as 'culture' is a contested concept, the justifications for a *sui generis* regime are likely to remain highly controversial, and potentially divisive. In situations where no distinct societal culture is identifiable, of course, there may be alternative legislative models dictated by policy – for example, the government holding designs on trust as the common heritage of the nation,¹⁴³ or forms of equitable benefit-sharing.

The corollary of communal interests arguments: groups as rights-holders?

Accepting that communal property rights in traditional designs seem justifiable, practical problems remain: who is to exercise such rights – and how, consistent with arguments thus far?

Nickel¹⁴⁴ highlights two potential deficiencies of groups as rights-holders pertinent to any consideration of the grant of proprietary rights to a cultural group: effective *agency* (the ability collectively to form goals, act deliberately, etc); and clear *identity* (ie group membership and boundaries). Both of these problems have been raised in arguments against granting property rights in intangible cultural items. But not all groups are deficient by these standards. Group agency and

138 Janke, n 4 above, Recommendation 18.3.

139 Naqvi's term: n 79 above, 720.

140 Puri, n 8 above, 300.

141 Naqvi, n 79 above, 721.

142 Such arguments might also provide policy reasons for lowering the 'damage threshold' – cf n 96 above.

143 Kuruk, n 18 above, 799ff.

144 J. Nickel, 'Group Agency and Group Rights' in Shapiro and Kymlicka, n 115 above, 235; similarly Rose, n 134 above, 140.

identity may exist by virtue of new or extant autonomous institutions:¹⁴⁵ a characteristic of many indigenous peoples. The extent to which agency and identity are *actually* features of particular indigenous groups is a matter for assessment and will vary.

First, we must be precise about which group we are talking about. In Australia, a significant proportion of the Aboriginal population lives in urban areas; only some live in communities on traditional lands, and there is no one coherent Aboriginal identity.¹⁴⁶ But we should not seek group agency at the broadest level. Given the variation between Aboriginal communities, smaller ownership groups are more likely to possess distinct societal cultures.

However, the coherence even of these communities is controversial.¹⁴⁷ Using concepts like agency and identity reflects the tendency to imagine the community as a 'person' with a coherent identity for the purposes of holding rights. 'Group as possessive individual'¹⁴⁸ is the dominant metaphor in world political cultures – making claims framed in these terms more likely to succeed.¹⁴⁹ But critics¹⁵⁰ argue against this tendency, for reasons already mentioned above, pointing out that the boundaries of cultures are fluid: individuals fall within communities to greater or lesser degrees; individuals within communities hold differing views; and there may be sub-groups within the community.¹⁵¹ Nor is there any guarantee that the boundaries of a political community correspond to the group claiming interests in a design.¹⁵² Can rights, or property interests, be based on such 'fuzzy' distinctions?¹⁵³

Admitting the impossibility of theoretical perfection, nevertheless, if protecting cultures is sufficiently important, then compromises may be needed to frame workable regimes. If control is to lie with the community, and unless we allow a complete overhaul of the dominant legal system, we must be able to identify a rights-holder capable of exercising rights. This is arguably even more important where the subject matter concerns intangibles. In practical terms, agency is more important than identity. Difficulties in defining the entire membership of a group should not be reason for denying the existence of a group proprietary right.¹⁵⁴ If rights can be exercised by some institution recognised both externally and internally as legitimate, then problems in identifying membership can also be managed through that institution. This would accord with the preference of Australian Aboriginal people; moreover, there are dangers of seeking to define a group from outside.¹⁵⁵ Internal relations may be governed by internal rules such as

145 Nickel, *ibid* 252–3.

146 P. O'Shane, 'Aborigines' future rests with empowerment' *Weekend Australian* 22–23 April 1995, 30.

147 For some it may not be possible: R.V. Annradha, 'In Search of Knowledge and Resources: Who Sows, Who Reaps?' (1997) 6 *Review of EC and International Environmental Law* 263, 270.

148 cf. M. Strathern, 'Potential property, intellectual rights and property in persons' (1996) 7(1) *Social Anthropology* 17, 23.

149 R. Handler, 'Who owns the past?' in B. Williams, *The Politics of Culture* (Washington DC: Smithsonian Institution Press, 1991), 63, 71.

150 Coombe, n 67 above, 254, 265; C. Severi, 'Comment' (1998) 39 *Current Anthropology* 214, 215; Descola, n 115 above, 209.

151 Kuruk, n 18 above, 769; J. McGowan and I. Udeinya, 'Collecting Traditional Medicines in Nigeria' in T. Greaves (ed), *Intellectual Property Rights for Indigenous Peoples: A Sourcebook* (Oklahoma: Society for Applied Anthropology, 1994) 59, 62.

152 Also n 18 above, 803; F. Granero, 'Comment' (1998) 39 *Current Anthropology* 214.

153 Pogge decries the idea: n 115 above.

154 Brennan J, *Mabo*, n 102 above, 61.

155 J. Cardinal-Schubert, 'In the Red' in B. Ziff and P. Rao, *Borrowed Power: Essays on Cultural Appropriation* (New Brunswick, NJ: Rutgers University Press, 1997) 122, 132; R. Paschal, 'The Imprimatur of Recognition' (1991) 66 *Wash LR* 209.

customary law,¹⁵⁶ without harming a group's capacity to be a rights-holder, provided the group has some mechanism for interacting with the broader society.¹⁵⁷ I do not underestimate the difficulty of identity questions, but the law has coped with such issues in the past;¹⁵⁸ on the other hand, agency is necessary before such decisions can even be mediated internally.

As for agency, one difficulty is determining what are legitimate institutions – democratic institutions in the familiar Western sense, or traditional (potentially non-democratic) institutions? In cases where suitable autonomous institutions do not already exist, a second difficulty is the apparent contradiction between creating institutions, and the aim of affirming the community's own culture.

In practice, states have either created new institutions, or defined relevant property-owning communities to include only those already given legal recognition – eg those already incorporated or possessing representative organisations.¹⁵⁹ Janke in her report for ATSIC outlines a range of different possible entities, and notably does not base her recommendations on institutions wholly traditional in form – proposing use of either existing land councils, or new regional and community-based organisations to exercise rights, with some central organisation to deal with inquiries and direct potential users to relevant groups, and a tribunal to mediate disputes – the emphasis being on the ownership rights remaining with and being controlled by the community.¹⁶⁰ Such approaches illustrate a pragmatic attitude towards the exercise of rights by the community.¹⁶¹

Are such institutions legitimate, according to the purposes considered earlier? This is questionable; however, once again we must beware of determining 'legitimacy' according to fixed notions of authenticity. If cultural integrity protects presently existing cultures, then institutions need not be the same as in pre-colonial times. Questions of cultural integrity depend on the distinct existence of a societal culture, which may – but will not *necessarily* – be threatened by introduced institutions. Legitimacy today should depend on community involvement in institution-building, and use of community norms and laws: for example, institutions that use kinship relations (eg *djungayi*) are more legitimate than a democratically elected land council only if preferred by group members. Individuals within communities, and those that speak for communities are political and social actors, with a role in the creation and recreation of communal institutions in changing circumstances, and perhaps in reaction to external influences. This is not necessarily an artificial process; such developments may even assist in strengthening underlying group identity. The additional danger that new bureaucracies will develop their own self-perpetuating logic – and overwhelm customary methods¹⁶² – is not confined to the arena of cultural control and traditional designs.

¹⁵⁶ But see 235 below.

¹⁵⁷ This is a 'limited common property' idea: Rose, n 134 above, 132.

¹⁵⁸ N. Lofgren, 'Defining Aboriginal Identity' (1995) 73(3) *Aboriginal Law Bulletin* 18.

¹⁵⁹ eg the Australian *Aboriginal and Torres Strait Islander Heritage Act* 1984 (Cth), which Golvan suggested might be extended to intangible property: n 72 above, 231. The Philippines adopted a similar approach regarding traditional knowledge: D. Daous, 'Efforts at protecting traditional knowledge: the experience of the Philippines' (1999) (WIPO/PTK/RT/99/6A).

¹⁶⁰ Janke, n 4 above, Recommendations 18.13, 18.15, 18.16.

¹⁶¹ D.A. Posey envisages the *community* taking responsibility to establish consensus on representation etc: 'International Agreements and Intellectual Property Right Protection for Indigenous Peoples' in Greaves, n 151 above, 225, App 1, Principle XII. This is less workable in the artistic context, where connections between indigenous group and user are less direct.

¹⁶² Brown, n 8 above, 204.

Further problems may arise where more than one community is associated with a particular design: for example, where a number of different Aboriginal communities have relationships with the same Creator Ancestor.¹⁶³ The only just solution is for all concerned groups to have rights in such designs; this is compatible with an aim of protecting cultural integrity for all Aboriginal communities equally. There is no reason to think that a *sui generis* regime could not be framed so as to cope with multiple interests, something copyright law already frequently does.

The interaction between group and individual interests

What solomonic process will sort out and create enlightened, sustainable policy upon the balance of rights among individuals, culturally identifiable collectivities, commercial interests, and the long-term public good? ¹⁶⁴

Thus far the focus has been the community's perspective, however in justifying property rights, the more important consideration is arguably what such rights *prohibit* people from doing.¹⁶⁵ Private property enhances one person's freedom at the expense of others;¹⁶⁶ communal property may enhance the power of a community acting communally, at the expense of both members' and non-members' liberties.

Interests liable to be affected include authors' copyright interests, and individual interests in free (artistic) expression. Three main categories of persons may be burdened:

- Non-indigenous individuals/corporations;
- Indigenous community members;
- Non-traditional indigenous artists.

Non-indigenous individuals

Inappropriate commercial use of traditional designs by non-Aboriginal people is, owing to both its economic and cultural effects, the most frequent source of complaint by Aboriginal groups.¹⁶⁷ Government papers¹⁶⁸ – and even critics of property rights in indigenous cultural intangibles¹⁶⁹ – tend to agree indigenous peoples are entitled at least to remuneration for the commercial exploitation of traditional designs. Most agree there should be a right to prevent such use; and certainly in the archetypal situation of mass-produced consumables, it is not unreasonable to require potential users to obtain permission. But there is no bright line between morally permissible and impermissible uses;¹⁷⁰ controversy surrounding less clear-cut cases is illustrated by the 'cultural appropriation'

¹⁶³ This may be in part a question of defining the subject matter.

¹⁶⁴ D. Tuzin, 'Comment' (1998) 39 *Current Anthropology* 217; Report of the Committee of Inquiry into Folklife in Australia, *Folklife – Our Living Heritage* (AGPS: 1987) 256 (referring to the intertwined rights of individuals and communities, and ethical and economic dimensions as a 'Gordian knot').

¹⁶⁵ Waldron, n 76 above, 842.

¹⁶⁶ Hettinger, n 129 above, 35.

¹⁶⁷ Janke, n 4 above, ch 3.1.

¹⁶⁸ Folklore Report, n 55 above, s 1219; Attorney-General's Department, 'Stopping the Rip-offs: Intellectual Property Protection for Aboriginal and Torres Strait Islander Peoples' (Australia: International Trade Law and Intellectual Property Branch, 1994) s 30.

¹⁶⁹ Brown, n 8 above, 204; L. Hiatt, 'Comment' (1998) 39 *Current Anthropology* 209.

¹⁷⁰ D. Vaver, 'Some Agnostic Observations on Intellectual Property' (1991) 6 IPJ 125, 128.

debate.¹⁷¹ Imagine a contemporary non-indigenous artist, expressing a political comment on contemporary Australia, in a way that incorporates the symbolic power of a traditional design.¹⁷² The wise law-maker, seeking a law in accordance with 'the prevailing sense of what is just'¹⁷³ here encounters a dilemma somewhat like that referred to by Strathern.¹⁷⁴

On one hand, freedom of artistic expression is important,¹⁷⁵ as is the encouragement of varied artistic productions. Symbols are a powerful means of expressing messages or meaning, potentially even hostile views, opinions, or attitudes.¹⁷⁶ An artist may plausibly argue that *only* the particular design suffices to convey the intended meaning, and may suffer genuine distress if prevented from using it.¹⁷⁷ Nevertheless, while artistic freedom and individual autonomy are potentially compromised, arguably, the artist's self-identification is not affected in the same way as that of group members. And the restriction on artistic liberty is limited: traditional designs are a small proportion of available sources when considered against the background of all possible contents of the public domain.

On the other hand, as argued above, the cultural harm to indigenous people that results from inappropriate use, especially of sacred designs, is undeniable. Furthermore, a community's traditional designs are limited in number, and of irreplaceable significance.¹⁷⁸ Such factors strongly suggest a different level and kind of harm to that suffered by the artist. If the arguments above are accepted, the continued vitality of the culture is threatened.

Further arguments may support restrictions even in this sympathetic situation. We are not necessarily concerned here with a complete restriction: in some cases, permission may be obtained, as Daes suggests.¹⁷⁹ It is interesting too that arguments usually asserted against *commercial* appropriation of cultural symbols – that those who deliberately disseminate a potent cultural symbol have a less just claim to exclusivity in symbols thus *imposed* on the culture¹⁸⁰ – do not apply with the same force to indigenous peoples' claims for control over these cultural symbols.

On balance, at least theoretically, it can be agreed that a *sui generis* system that prohibits use of traditional designs, without prior authorisation, by non-indigenous individuals and corporations in all circumstances is justifiable, most importantly because the interests involved for indigenous communities in these circumstances are of a different order to those at stake for non-indigenous individuals. Individual artists do have claims in these situations, and individual injustices will occur, but this is inevitable in framing a workable law.¹⁸¹ Provided that my earlier arguments

171 Paak, n 10 above, 58–62; Coombe, n 67 above.

172 This is not uncommon: Gray, n 117 above, 30–31.

173 W. Cornish, *Intellectual Property* (London: Sweet & Maxwell, 4th ed, 1999) ss 1–12.

174 Strathern, n 1 above.

175 For a rejection of free expression arguments: Daes, n 8 above, ss 24–26.

176 W. Gordon, 'A property right in self-expression' (1993) 102 YLJ 1533, 1568–1569; Coombe, n 14 above, 480; for an example, see the 'parody' cases, eg *Walt Disney Prods v Air Pirates*, 581 F.2d 751 (1978).

177 J. Boggs, 'Who Owns This?' (1993) 68 Chi-Kent LR 889; Spence, n 127 above, fn 43.

178 Space prevents me discussing the varying significance of designs and perhaps the varying degrees of protection justified. I suspect the attempt to draw such distinctions, even if theoretically ideal, would be unworkable.

179 Daes, n 8 above, Guideline 46. Whether this is realistic is questionable: cf the description of the artistic process in Boggs, n 177 above.

180 Waldron, n 76 above, 885; R. Coombe, 'Objects of Property and Subjects of Politics' (1991) 69 Tex LR 1853, 1872.

181 Janke, n 4 above, suggests there should be no innocent infringement exception (recommendation 18.12); however, framed restrictively, it might benefit individual artists.

above are sound, the balance of interests in most cases falls in favour of the community.

Community members

Two issues arise in this context. First, would a *sui generis* regime restrict the entire community to customary uses?¹⁸² Ellinson suggests consistency might require such a result¹⁸³ – but by precluding innovation, this could stymie the development and vitality of the culture. However, customary law regarding the use of traditional designs can and has changed over time.¹⁸⁴ I have argued that cultural integrity does not preclude such development; the state is not justified in interfering in the community's decisions on which traditions to keep, discard, or develop. For so long as designs play an integral role in the culture,¹⁸⁵ protection is justifiable.

More interesting issues relate to the allocation of power in community life, especially group control over individual action. The unitary action implicit in group agency means that granting group property rights in cultural expression will limit freedom of artistic expression *internally*. We must therefore consider the impact on the 'dissenting member', whose particular use of designs is not approved by the community acting through its institutions ('the community') either because the proposed use is non-customary, or because the dissenter is not entitled to use the designs at all.

The grant of group control over forms of expression sits uncomfortably within the framework of liberal individualism central to the Western democratic political system.¹⁸⁶ Criticisms of group rights often focus on potential restrictions to individual liberty, and this area is no exception – some critics have predicted dire consequences:

By conferring upon an ethnic nation the right to suppress ideas and productions deemed to be offensive to its subjects, it would in fact be equipping dominant factions with a legal mechanism for discouraging dissidence and silencing rivals.¹⁸⁷

Prior to creation of a work, this is probably a matter for the community and its social norms. The question for the dominant legal system arises once a work of art incorporating traditional designs has been created. Because copyright subsists in at least some, if not most works embodying traditional designs, an important question is whether continued existence of *individual* interests is appropriate, or whether works embodying such designs are to be excluded from copyright. This is an open question in the literature.¹⁸⁸ Golvan proposed that new rights would sit alongside the copyright system and continue after the copyright period ended.¹⁸⁹ Janke suggests that *sui generis* legislation would not override existing copyright, but would apply outside the copyright period or where copyright did not subsist.¹⁹⁰ The important question is one of priority, where the dissenting member wishes to use

182 Customary use would be allowed: *ibid*, recommendations 18.6, 18.12; Folklore Report, n 55 above, s 1227.

183 Ellinson, n 11 above, 340–341.

184 Toobey, n 41 above, 4.

185 Different questions might arise where designs no longer play such a role.

186 R.N. Clinton, 'The Rights of Indigenous Peoples as Collective Group Rights' (1990) 32 Ariz LR 739, 740–741; Pask, n 10 above, 85. For the international debate regarding collective rights: ATSIC, *An Analysis of the UN Draft Declaration on the Rights of Indigenous Peoples* (1999) (www.atninc.gov.au).

187 Hiatt, n 169 above; also Brown, n 105 above, 218.

188 Ellinson reaches no firm conclusion: n 11 above, 343.

189 Golvan, n 72 above, 230–231.

190 Janke, n 4 above, s 18.10; recommendation 18.7; *Mataatua*, n 7 above, Principle 2.5.

copyright inconsistently with the community's dictates. What is the role of a *sui generis* law here?

First, the law could intervene to enforce communal interests. In effect,¹⁹¹ the community would have a right of consultation and veto. This was effectively the result in *Bulun Bulun*, where the Court found the artist was under a fiduciary duty not to use his copyright in ways contrary to customary law.¹⁹² Second, the law could abstain from enforcing communal interests against the individual, and so allow the artist to use their copyright even in conflict with the group's communal interests, subject only to communal mechanisms. Theoretically, the law could 'stay out' of the dispute – however, a limbo where neither copyright nor *sui generis* rights were enforceable is inconceivable as a practical matter.

Is there a principled basis for resolving these conflicts? A utilitarian approach is unhelpful – it is not sufficient to argue that the benefit of the many outweighs the sacrifice of individual interests.¹⁹³ Cultural integrity is not justified by strength of numbers; moreover, such a solution violates the Kantian injunction against using individuals as means rather than ends: we must consider the relative importance of the individual and group interests involved.¹⁹⁴ Nor is it sufficient to say that the individual has accepted membership on terms that rights should be exercised collectively;¹⁹⁵ by the time the individual is old enough to decide whether to accept membership, designs and expressions of that culture are already part of their identity. Another argument is that restrictions on the freedom of individual members are justified, provided the individual has the right to leave.¹⁹⁶ This too is unconvincing: the social dislocation involved in leaving may render the 'right' illusory (especially given the importance, mentioned above, of one's *own* culture)¹⁹⁷ and the *sui generis* regime would burden the individual even after departure.

An alternative approach may be to consider the underlying purposes of the competing group and individual rights, and the relationships ('internal relations') between such purposes, to determine whether in the context of this competition, one cannot be asserted consistently with those underlying purposes.¹⁹⁸ The situation of the dissenting community member involves a clash of identities: individual identity and self-expression as against the communal integrity and collective identity considered earlier. One justification of cultural integrity is the role of societal cultures in promoting individual autonomy, and facilitating individual choice. But free artistic expression is *also* an important element of individual autonomy – and arguably necessary if the societal culture is to provide a context for individual choice. The individual's interest in using the designs is even more powerful here than at p 234 above: in this case, the symbols involved are part of the individual's *own* culture *and* identity. If we accept that a purpose of both rights is to promote individual autonomy, then in this particular case, where

191 For present purposes, I set to one side the interests of third parties.

192 Janke's proposal could have the same effect if *Bulun Bulun*, n 26 above, remains good law, depending on the facts (*ibid* 209).

193 McDonald, n 88 above, 324.

194 J. Waldron, 'Rights in Conflict' in Waldron, n 104 above, 203, 210–212.

195 J. Triggs, 'The Rights of "Peoples" and Individual Rights: Conflict or Harmony?' in Crawford, n 121 above, 141, 149.

196 C. Kukathas; 'Are there any cultural rights?' in Kymlicka, n 79 above, 228, 238.

197 Y. Mokgoro, 'The Protection of Cultural Identity in the Constitution and the Creation of National Unity in South Africa: A Contradiction in Terms?' (1999) 52 SMULR 1549; L. Green, 'Internal Minorities and their Rights' in Kymlicka, *ibid* 257, 264.

198 Waldron, n 194 above, 220 ff; McDonald, n 88 above, 328–9.

individual autonomy is limited in an important respect by protecting cultural integrity, arguably the individual interest must prevail.¹⁹⁹

Such an approach, which subordinates group concerns to the interests of any individual member, essentially ignores any collective aspect to the issues involved, ignores interests of other members of the group in the culture fundamental to their self-identification, and could obliterate the purpose of group rights. To the extent that culture is a good that cannot be enjoyed alone, the assertion of autonomy in disregard of group norms by an individual may frustrate the whole purpose of cultural integrity. Furthermore, the individual artist using traditional designs as a personal (dissenting) expression may in some fundamental way deny other members the opportunity to pursue their (communal) interests. For while traditional designs are non-excludable, their power and significance to the group is diminished by misuse.

The complex of different interests involved (even ignoring economic interests) and the uncertainty surrounding the true purposes of cultural integrity (in particular, its group or individual nature) indicate that an internal relations approach will not provide an answer. A further approach to balancing conflicting rights is to assess the relative importance of the underlying interests. But even here we encounter complications: in so far as different views exist as to the relative importance of group and individual interests, balancing is impossible. Furthermore, if you take the extreme case, where the continued existence of the culture is at stake if restrictions are not imposed on the individual's freedom to act, and if you say that in that case restrictions are justified, you are asserting that the state will intervene to prevent (artistic) revolution, or rather, will intervene to preserve a community's preferred version of itself.

The attempt to balance interests is further complicated by certain imponderables: it is difficult, if not impossible to determine the extent to which control of designs is *necessary* to the continued survival of the distinct culture, as opposed to merely *assisting* that aim (how can we know ahead of time?); or whether perhaps only certain particularly significant designs are necessary in this sense (how are distinctions to be drawn?); or the extent to which the survival of the culture will benefit from being open to challenges (often argued, but unprovable). The difficulties of grappling with these issues in individual cases are multiplied by the need, in devising a legislative regime, to make some choices in advance. A regime based on consideration of how important the individual's claim to free expression was in a particular case would be unworkable; approximations and compromises therefore are necessary, but particularly difficult to assess in advance.

Given these difficulties, no wonder it is frequently asserted that western law should not intervene in non-traditional use of designs by community members, but should leave such matters to communal dispute mechanisms.²⁰⁰ This approach must be preferable wherever possible. But what is to be the relation between the *sui generis* system and copyright, and what is the court to do if asked to adjudicate? What of situations where third parties are involved, for example as copyright licensee?

Earlier I characterised the choice between prioritising group or individual interests as a political choice; as a result, the attitude one takes to this question also depends on a political choice between individual and group interests. My own (preliminary) view is that in principle, where no third party interests are involved,

¹⁹⁹ This is consistent with Kymlicka's prohibition on 'internal restrictions': n 87 above, 152.

²⁰⁰ Janke, n 4 above, § 18.8; Coombe, n 67 above, 279; Gray, n 73 above, 35.

the Court should not be involved. If the underlying purpose is protecting the distinct existence of the culture, including its institutions and customary law, and if those community mechanisms are not sufficient to resolve an internal dispute, then the community should not then be able to call on state power to enforce its internal rules on the individual member.²⁰¹ An exception might apply in some cases: the law might refrain from intervening to assert a community's interests where individual artworks by a group member were involved, but intervene to prevent use of copyright by the individual artist for commercialisation, especially on a mass scale, based on the judgement that such use is more damaging to the culture and less justified by the individual's interest in free expression.²⁰² Similar reasoning might apply to cases of extreme offence to community beliefs – but that would be closer to a 'right against change' or 'right to be protected from criticism' and hence more problematic. Where third-party interests are involved, some court involvement may be inevitable. One possibility is that the *sui generis* system could provide for compulsory consultation with the community as well as the individual prior to use – so that potential difficulties within the community can be considered in advance. Court enforcement in the absence of consultation could be denied. This would restrict the freedom of the individual artist to deal with his or her work, but would at least air issues in advance of any direct involvement of third-party interests.

The reader will appreciate that these suggestions favour individual interests where community solutions fail, and the view that state power cannot legitimately be used to suppress internal dissent – even artistic dissent. The argument is based on the view that cultural integrity is not enough to justify the suppression of individual autonomy and identity of members. Arguments against this approach would need to explain why the collective identity of the community is able to prevail over the individual interest in autonomy in self-identification and expression.

Non-traditional (urban) indigenous artists

Urban indigenous artists are not members of traditional communities, and hence are not bound by customary laws, nor do they have customary entitlements to use traditional designs. But they are also not in the same position as non-indigenous individuals, precisely because cultures shade into one another, and because cultural heritage and identity are personal as well as cultural concepts. Gray²⁰³ has considered the position of the urban indigenous artist, who may need to draw on traditional artforms in expressing their identity by reference to their Aboriginal heritage:

You have to understand my position of having no designs or images or stories on which to draw to assert my Aboriginality ... Dots are my bridge to my Aboriginality. They connect to their obvious relationship to the Western Desert paintings and their relationship to reproductions and representations ... I am continually searching for ways to express my Aboriginality ...²⁰⁴

201 Different considerations could apply if (a) community mechanisms were oppressively harsh, and/or (b) community members were prevented from leaving. Use of traditional designs does not involve such extreme restrictions; compared, for example, with issues relating to female genital circumcision.

202 This approach would also take economic considerations at 228 ff above into account. This distinction between commercial use and 'high' art is justified by the different balance of interests in this case; cf 234 above.

203 Gray, n 73 above.

204 Gordon Bennett (an artist of Aboriginal descent), quoted, *ibid* 31. See also Ellinson, n 11 above, 341.

Traditional designs here play a different role: not as expressions of communal culture, but as articulations of political consciousness and identity. There is a strong basis for the claim that use of such designs is fundamental to these artists' identity and autonomy.

These are questions with significant practical consequences. Indigenous people are concerned about the use of traditional designs by non-traditional artists;²⁰⁵ even the major Aboriginal arts funding body has expressed disquiet.²⁰⁶ Such issues will only become more pressing, as the Aboriginal art market becomes more valuable, and as urban Aboriginal art gains popularity.²⁰⁷ Again, there are significant issues regarding the power of the community, and the relationships between the purposes of cultural integrity and the claims of an individual whose liberty is restricted. I focus below on features that differentiate the situation of the urban indigenous artist from the dissenting member.

First, communal mechanisms are unlikely to be sufficient to resolve disputes, and we cannot argue that, if the community's own mechanisms fail to resolve the issue the community is not entitled to seek state power in support – the state provides the only common *legal* system. There are also difficulties of principle with recognising these claims. By compromising the interests of the community which lives as a distinct societal culture, acceptance may expose cultural integrity to the criticism that it is, at core, a claim of ownership through peculiar attachment, that can be matched or overridden by opposed claims of attachment unrelated to any societal culture. The difficulties of drawing lines between individual cultural heritage and group cultures is exposed. Racial discrimination is also a problem: what other logical distinction can be drawn between an urban indigenous artist of mixed descent, and an artist brought up under the pervasive influence of Aboriginal culture but with no racial ties – are not such designs equally important to the latter individual's identity? And if so, can you go further and acknowledge the claim of an urban, non-indigenous artist for whom the influence of Aboriginal art has had a deep, personal effect that has shaped their perception of self? Furthermore, do we draw a distinction between urban indigenous *artists* and urban indigenous people with an eye for a commercial chance – and how? The distinction in principle lies in the importance of use of designs in self-fulfillment and the expression of identity²⁰⁸ – but how is this to form the basis of a legal distinction?

The literature offers no satisfactory solutions. Ellinson again points out that consistency requires that the law prevent use by urban indigenous artists, because the interests of the traditional community are no less affected by such use.²⁰⁹ Janke's report is notable for raising the issue, but suggesting only that legislation 'contain special provisions to protect the artistic heritage and cultural products of urban Indigenous artists' – without explaining what 'heritage' means or what relation it bears to traditional artforms.²¹⁰ Gray²¹¹ argues for consultation with urban indigenous artists in framing laws.

²⁰⁵ Janke, n 4 above, s 3.15.2; Ellinson, *ibid*.

²⁰⁶ Gray, n 73 above, 32.

²⁰⁷ *ibid* 33, 35.

²⁰⁸ This may suggest an implicit hierarchy prioritising individual identity over economic concerns. There is no space to enter here into a general discussion regarding prioritising of rights and interests.

²⁰⁹ Ellinson, n 11 above, 341.

²¹⁰ Janke, n 4 above, s 18.4.7.

²¹¹ Gray, n 73 above, 42.

Strictly applied, a cultural integrity rationale requires that urban indigenous artists be prevented from using traditional designs without permission.²¹² In practice, the answer cannot be so categorical. No rationale could require protection at all costs, and provided the whole purpose is not thwarted, countervailing interests may sometimes prevail.²¹³ Applying the arguments explored at p 235 above leads to recognition of the significant role that concerns of autonomy and identity must play in the calculus, and the role of art in such identity. Fundamental interests in identity are involved here for the urban artist as well as the group. We may conclude that Aboriginal communities cannot claim *automatic* precedence for the autonomy and identity of their members over the search for identity of urban indigenous artists.

I am led again to the conclusion that informal mechanisms, eg mediation and consultation between urban artists and traditional communities are to be preferred, but that, as a last resort, the law should not impose prohibitions on use of traditional designs by urban indigenous artists. Respect for cultural integrity issues could be encouraged by other means: for example, through education, facilitating consultation, or making consultation or non-use of traditional designs a condition of funding. This conclusion, despite its problems, recognises the importance of the claims to individual identity involved for the urban indigenous artists, which are no less significant than similar claims by community members. The contrary position is also likely to be interpreted as meaning that the only vision that the 'colonial' law considers worthwhile is traditional;²¹⁴ and risks creating classes and tension within the Aboriginal population. A final, perhaps obvious reason in favour of this solution is the general public interest in encouraging urban Aboriginal art – a generally well regarded art form.

In recognition of the explicit compromise being made between competing interests, some restrictions could be placed on use by urban indigenous artists: again mass commercialisation, or use of sacred-secret designs could perhaps be prohibited in recognition of the greater harm such use causes. *Essential* limits on use by non-members of communities (and more firm definitions of the limits suggested above) could perhaps be the subject of consultation in the drafting of *sui generis* legislation.

Conclusions

The claim to communal ownership of traditional designs by indigenous people is inextricably tied to concerns to protect cultures of indigenous peoples – cultures which are under threat. Cultural integrity, or control over culture, has become one of the key claims of indigenous peoples in their international efforts to gain recognition and accommodation of their interests in legal systems that have, in the past, ignored or excluded those interests. The political nature of such claims, and the potential of such claims to impact on a multiplicity of other interests and the liberties of many non-indigenous as well as many indigenous persons, makes it very important to articulate the justifications for such claims, and the values that they serve.

212 Ellinson, n 11 above, 341.

213 Coombe, n 8 above, 86.

214 See similar criticisms in Coombe, n 67 above, 268.

In this article I have sought to outline some of the assumptions and interests that underlie claims to traditional designs, in order to consider more closely how such interests may be reflected in rights capable of being exercised, and importantly, how such interests interact with individual freedom of artistic expression. It can be seen that while a principled argument can be made for such group interests, once they *are* recognised, complex and controversial decisions as to priority need ultimately to be made between individual and collective interests – issues that the same principled arguments will not necessarily resolve. The inseparable and contested nature of group and individual interests in culture make the effort to find laws that equate with our sense of justice exceedingly difficult.

Further, a fundamental uncertainty lies at the heart of the claim for cultural integrity, owing to the uncertain, fluid nature of culture in a increasingly interconnected world. The theoretical difficulties that this poses defy logical resolution; the practical implications may be more solvable, by seeking approximations or analogues to 'societal culture', in those instances (in those countries, by those groups) where a supportable claim is made. Even if such compromises are found, the uncertainty regarding underlying concepts will be a source of ongoing controversy.

Many questions remain. First, I have only considered traditional designs. Given the emphasis on cultural integrity, self-expression and identity, it appears that the arguments considered above will apply more directly to other artistic or cultural subject matters, but not so directly to questions relating to traditional knowledge. I suspect that economic considerations would figure more prominently in relation to traditional knowledge, although, even in that case, questions of spirituality are relevant (given the holistic world-view indigenous peoples adopt, and the 'cultural' implications of matters Western views would consider scientific), and questions of control and interference are likely to arise (for example, in relation to control of researchers interacting with indigenous people).

One major issue, which arises from the framework suggested here, and which I have not addressed, is how these same principles might affect the definition of rights to be attached to the designs (or other indigenous cultural and intellectual property), and specifically, the role of customary law in defining such rights. A key question in this context is whether rights are to be defined according to customary law values and ideas of appropriate use and treatment of designs, as opposed to more traditionally Western views. Indigenous advocates consistently emphasise the importance of treating indigenous intangible property in accordance with indigenous world-views and values. This would seem to be supported by an approach that focuses on protecting the integrity of the culture as a distinct whole. However, I have focused on the prior issue of control; the debate about the nature of the rights and interests to be asserted is a further, equally controversial question.

Some might say that the traditional designs of the Australian Aboriginal people represents the 'easy case' in assessing the interaction between communal and individual interests in valuable intangibles. They are, after all, an autochthonous people, clearly identifiable and possessing a distinct culture, as well as very distinctive traditional artforms in at least some cases associated with particular, relatively self-contained communities. They have greater access than many indigenous peoples to legal resources and so a greater capacity to articulate their claims both in terms of the dominant Australian legal system, as well as in their own terms. I have not had to consider the further complications that may arise where, for example, indigenous peoples straddle national borders or where survival

LEGISLATION

New Labour's Welfare Reforms: Anything New?

*Julian Fulbrook**

The Labour Government's path to legislative reform of the welfare system has been characterised by internal party dissension, a mauling from both Opposition parties, and repeated defeats in the House of Lords. Two major statutes have been enacted, but with the promise of further modernising reforms to come, tactical considerations in the run-up to a General Election suggest that there may be a lull before the next legislative combat. In the House of Commons the 1999 Welfare Reform and Pensions Bill suffered repeated rebellions from Labour backbenchers, and even though the scale of the revolt declined steadily, from 67 in May to 54 in the final vote in November 1999, risking further difficulties might be considered unwise in the short term of the electoral cycle.

Despite many protests, and a background of controversy over the amount of the state retirement pension, the Government refused to back away, and few concessions were made. Having survived one serious battle on welfare reform in 1999, the Government returned for a second round with another controversial set of measures in 2000. Taken together these statutes have made important changes to the social security system, but how far can they be described as a 'New Beveridge?' Both the Welfare Reform and Pensions Act 1999 and the Child Support, Pensions and Social Security Act 2000 are now on the statute book, so it is perhaps a useful moment to reflect on these measures and also to look at their historical context.

Alistair Darling, the Secretary of State for Social Security, has repeatedly indicated how welfare reform is central to the Government's whole modernising programme. The aim has been to 'modernise the welfare state so that it meets the modern needs of our country ... I believe our reforms, based on principle and recognising the changes in our society over the past 50 years, will help lay the foundations for a modern welfare state in the 21st century.'¹ In this endeavour, Mr Darling has obviously been following the line promoted by both No. 10 and No. 11 Downing Street. However, the Government's 'big agenda for change and reform' has certainly had a rough passage so far. The Prime Minister has repeatedly declared that 'Welfare reform goes to the heart of what Labour stands for. It is the issue on which our government will be judged'.²

* Law Department, London School of Economics and Political Science. I am grateful to several colleagues for their comments on an earlier draft of this article, and in particular to Professor Nick Wikeley of Southampton University, Professor Adrian Sinfield of Edinburgh University, and Dr. Peter Wright, Office of the Chief Medical Adviser, DSS.

1 *The Guardian* 5 November 1999. See for a recent assessment of Mr Darling, Jackie Ashley, 'The NS Interview', *New Statesman* 2 October 2000.

2 See in particular the Beveridge Lecture at Toynbee Hall in March 1999, reported in *The Times*, 19 March 1999.

Thinking the unthinkable

When, a year after the General Election, little seemed to have been achieved, other than the publication of a Green Paper in March 1998 entitled *New Ambitions for our Country: A New Contract for Welfare*,³ the Prime Minister abruptly changed his social security team. Harriet Harman, the then Secretary of State and repeatedly under opposition fire for a 'wasted year' was sacked. Her deputy, the long-term welfare campaigner Frank Field, chose the more maverick route of resigning as Minister for Welfare Reform, rather than take a sideways move to oversee the problem of fraud. Most of the rest of the junior team, apart from Baroness Hollis (who still remains), were moved to other departments.

The travails of Frank Field, now 'Fearless Frank' to the readership of mass circulation newspapers where he has carved out a new career, is a highly illustrative case study of the Government's difficulties and philosophical approach to these vexed matters. The original Green Paper version of *New Ambitions* had been very much Mr Field's own work, although later drafts were filleted in Downing Street and the Treasury. Somewhat unusually Mr Field, as the number two in the Department, introduced the final version to the House of Commons.⁴ It offered a 'new' drive against benefit fraud, action to get claimants into jobs and the first steps towards compulsory saving for second pensions. The Government promise was to break the 'dependency culture', as part of the 'New Contract' with the electorate. In a personal foreword, the Prime Minister promised 'work for those who can and security for those who cannot'; this remains the central tenet. Mr Field, a former director of the Child Poverty Action Group and of the Low Pay Unit, claimed in his speech that the case for welfare reform was overwhelming. The first objective of the system would be to 'get people back into work' rather than simply paying benefits. The Government then set itself 32 'success measures' against which the reforms should be judged. They range from an increase in the number of working-age people in work to a fall in the number of pregnancies among girls under 16. However, these have not been given a time-scale, and the Government indicated that the changes should be judged 'over the next 10 to 20 years'. Iain Duncan Smith MP, the then Conservative social security spokesman, also promising 'Draconian' measures to root out fraud, described the Green Paper as 'vacuous' and 'a series of missed opportunities'.

Unfortunately, Mr Field, despite an illustrious thirty years of campaigning on welfare issues, had very little time left to make Government policy. He had been asked personally by the Prime Minister to 'think the unthinkable', but it may be that there was rather too much cerebral cogitation and not enough action. In particular, the Minister had been asked to cut the cost of welfare to the taxpayers and to root out fraud; but there had not been much of a practical outcome in the time-scale allowed. Mr Field had a spectacular few weeks in the spotlight following his resignation, and has now returned to the role of pundit. Perhaps teamwork was also an issue; he had been installed in the DSS to work in tandem with Harriet Harman, but claimed that 'it was like grafting someone else's organ on to a body. Chances of rejection are pretty great'.⁵ In the introduction to a

3 See the White Paper Cm 3805 (1998). See also the companion, *A New Contract for Welfare: Support for Disabled People* Cm 4103 (1998).

4 HC Deb vol 309 col 682 26 March 1998.

5 *The People*, 2 August 1998: 'Whatever has gone before, I won't give up. I will be extremely vocal. The benefits culture in this country must be changed. It rots people, it destroys their sense of responsibility to their family, it kills their will to work, it pays for them to be dishonest. It's also a

collection of essays, *The State of Dependency*, Mr Field stated: 'New Labour came to power committed to transform welfare. That revolutionary intent has fractured.'⁶ Now one of the most vocal critics of the Government, he points out that the values of work, saving and honesty were to have been at the core of the new economic and social order, and the welfare system was to be remodelled to help promote those values. Instead, Mr Field has claimed that

Welfare cuts have been pushed through Parliament against the most persistent opposition any government has experienced since the back-bench revolt brought down the Chamberlain government in 1940 ... Welfare changes, masquerading as reform, have undermined the principle of work, mocked the idea of saving, and weighted public policy against people who tell the truth.⁷

He also recently added to Mr Blair's woes at the hands of some slow-handclapping members of the Women's Institute by accusing the Government of 'being more interested in spinning to the media than delivering the improved public services demanded by the voters.'⁸

The Government has therefore had to weather serious backbench revolts, as well as many difficult private discussions with disaffected party members; the long list includes luminaries such as former Secretary of State Barbara Castle and former TGWU General Secretary Jack Jones. Major fault lines in the Labour Party were disclosed in the revolt over the reduction in Lone Parent Benefit in November 1997, and then the serious schism in May 1999, when a total of 67 Labour MPs voted against the Welfare Reform Bill clause which proposed to cut Incapacity Benefit for new claimants and restrict entitlement to those who had not paid sufficient National Insurance contributions. Since then, the Secretary of State, Alistair Darling, has been rather more successful in damage limitation on the main issues of reform legislation, although he has certainly had difficulties, along with other senior ministers, in explaining the level of the state retirement pension.

The Government welfare programme

For a government trumpeting the brand image of 'new', it is inevitable that judgements have to be made on what this means in social security. The history of welfare is littered with polemical debate, loudly heralded reforms which proved to be gimmicks, principle shown up as expediency, and overall a general lack of finance to do something positive for the poorest in the community. From a legal perspective, social security law is, as Professor Sir Otto Kahn-Freund famously remarked, 'not only complex [but] extraordinarily kaleidoscopic, so that everything said today may very well have lost its validity tomorrow'.⁹ Has the New Labour regime merely shaken the kaleidoscope, or is there a positive new direction? Certainly there has been a considerable contrast with the rhetoric of the Labour Government elected in 1997, when compared with the previous

moral issue because the benefits system affects how people behave. Would we have so many single mothers if the system didn't actively assist them in their situation?

6 See also Frank Field, *Making Welfare Work*, Institute of Community Studies, 1995, described by David Walker as 'rhetorical overkill, of the kind favoured by the messianic Frank Field, former minister for welfare reform. Field's acquaintance with abuses in Birkenhead was real enough, but talk of "SAS-style anti-fraud" squads is over the top.' *Analysis in The Guardian* 29 September 1998.

7 *Birmingham Post*, 27 March 2000.

8 *The Independent*, 9 June 2000, 1.

9 Otto Kahn-Freund in E. Stein and T.L. Nicholson (eds), *American Enterprise in the European Community Market* (Ann Arbor, Michigan: University of Michigan Law School, 1960) 298.

Conservative administrations stretching back to 1979. The leading campaign organisation on behalf of the poor, the Child Poverty Action Group (CPAG), summed up social policy evidence in their pre-election guide as 'a huge indictment of the disastrous record of four Conservative governments towards Britain's most vulnerable and insecure citizens'.¹⁰ CPAG pointed out that the social security system had 'been the subject of almost constant review and reform for over 20 years,' so there is perhaps nothing new in a government with reforming zeal in this area.¹¹ However, CPAG's verdict on Labour's manifesto was, perhaps predictably, that it was rather 'modest'.

Taking the long historical sweep, continuities in welfare have also been formidable. Despite the constant upheavals, there are many persistent themes. Perhaps the best text to analyse on New Labour's welfare reform programme is the Prime Minister's Beveridge lecture in March 1999 given at Toynbee Hall. This practically sacred location in East London was the first University Settlement, founded by Canon Samuel Barnett, an early proponent of old age pensions. Toynbee Hall had not only nurtured Beveridge's interest in social policy, but also that of other notables such as Clement Attlee, Richard Tawney and Hubert Llewellyn-Smith (co-architect of the 1911 National Insurance Act with Beveridge). Canon Barnett had also influenced, among others, Charles Booth, often described as 'the father of scientific social surveys', and his niece Beatrice Webb.

Amounting almost to a pilgrimage, the Prime Minister's Beveridge lecture rehearses his phrases of 'an active, not a passive welfare state and a system that is genuinely providing people with a hand-up, not a hand-out,' which he has now used on many occasions. Translating this philosophy into legislative practice has been a little difficult, particularly when New Labour determined that the previous government's financial planning should not be interfered with in the short term. This inevitably meant welfare cuts. As pointed out by the *Daily Telegraph*, 'For a party that made the heartlessness of its opponents into an Article of faith, the prospect of removing benefits from the disabled and sick must seem distinctly unpalatable.'¹² With expenditure on benefits for the long-term sick and disabled accounting for more than a quarter of the £100 billion spent annually on social security, and growing at a rate of 14 per cent a year in real terms between 1992 and 1996, it was not altogether surprising that the government of John Major had felt the need to make cuts in this area.

The defence used by New Labour is that it carried on this momentum for the short term, until the country's spending was in control. In particular, the Disability Living Allowance, introduced in 1992, accounted for a large part of the growth. William Hague's own period as a social security minister had seen the (much-televised) introduction of a 20-page questionnaire and new tests on Incapacity Benefit, for those who were thought to be able to obtain 'sick notes' from their own doctors rather too readily.¹³ Mr Hague inherited the policy, and the regulations were passed when he was in post, but perhaps his influence was fairly minimal. However, he was the minister who made the confident prediction that up to 250,000 people would be deemed fit to work and taken off benefit as a result of the tests. However, the impact of the Benefit Agency doctors, and the subsequent

10 A. and C. Walker (eds), *Britain Divided: The growth of social exclusion in the 1980s and 1990s* (London: Child Poverty Action Group, 1997) Foreword vi.

11 *ibid* 285.

12 *Daily Telegraph*, 16 December 1997.

13 For an excellent review of the debate on the 'reforms' of both governments, see G. McKeever, 'Welfare to Work for the (In)Capacitated - The Reform of Incapacity Benefit' (2000) 29 *Industrial Law Journal* 145.

tribunal appeals, amounted perhaps to rather less of a saving. Both the numbers transferring on to Incapacity Benefit, and the numbers on the new benefit, have fallen since it replaced Invalidity Benefit. But the relationship between cause and effect is not entirely clear, as other changes also limited payments.

A critical focus in New Labour's programme has been that the Government's top priority is 'children, one third of whom – around 700,000 – live in "frightening" deprivation'. With an audience of experts at Toynbee Hall, Mr Blair dealt with themes ranging from truancy and youth crime to teenage pregnancy and low skills, detailing the links between poverty and failure. Then came the pledge: 'our historic aim will be for ours to be the first generation to end child poverty. It will take a generation. It is a 20-year mission, but I believe it can be done.' The Prime Minister then listed Labour's reforms so far, taking a wide view of 'welfare': student finance; the New Deal welfare-to-work programme for 18–25s; disability reforms; the framework for pension reforms and minimum income guarantee; child support reform; and the working families tax credit scheme, with increased child benefit.¹⁴ Public support has also been an issue: 'The only road to a popular welfare state is radical welfare reform', Mr Blair concluded.¹⁵

Underpinning the public statements of the Government has been the work of the Prime Minister's Policy Unit, and textual analysis is useful here too. In a revealing lecture given to the British Association for the Advancement of Science, David Miliband, the Head of the No. 10 Policy Unit, indicated that the 'mission' of the Government was to 'modernise the institutions of the country to put power, wealth and opportunity in the hands of the many, not the few. The welfare state is at the heart of that project.'¹⁶ The difficulty with such a project is that the continuities in welfare show that many of the same questions have recurred in public debate over centuries. It is therefore unsurprising that the present Government has continued to agonise over intractable problems high on any social agenda.

Key themes: rough sleepers, beggars, and fraud

Perhaps the earliest piece of welfare legislation enacted by Parliament was the 1350 Statute of Labourers. It was 'anti-panhandling' law. In an attempt to re-impose the feudal order, Parliament empowered the justices to bind wandering labourers at their place of settlement. Their wages and conditions of employment were then to be arbitrarily fixed at pre-plague rates. The 1350 Act defined its subject matter as 'sturdy vagabonds'.¹⁷ The contemporary parallels are everywhere. Take, for example, the Government's Rough Sleepers' Initiative to sweep the streets clear of the homeless. The statement by Louise Casey, the Director of the Rough Sleepers' Unit, that 'With soup runs and other kinds of charity help, well-meaning people are spending money servicing the problem on the streets and keeping it there',¹⁸ could have come straight from the Victorian debate between the Charity Organisation

¹⁴ *The Guardian*, 19 March 1999.

¹⁵ *The Times*, 19 March 1999. For the full text of the Prime Minister's speech, and an analysis see R. Walker (ed), *Ending Child Poverty* (Bristol: Policy Press, 1999).

¹⁶ 'This is the modern world; the future of the welfare state' (1999) 111(4) *Fabian Review* 11.

¹⁷ 25 Edw. III, c. 2 (1350–1351). Pre-Parliamentary laws date from the reign of King Ethelred.

¹⁸ Quoted in *The Independent*, 15 November 1999. Ms Casey, the former Deputy Head of Shelter, is implementing plans to help rough sleepers, with the aim of reducing their number by two-thirds by 2002. DETR set up the Rough Sleepers Unit (RSU) on 1 April 1999. The RSU's remit was originally London, but has now been expanded to develop and co-ordinate a policy to prevent rough sleeping throughout England.

Society (COS) and the Salvation Army.¹⁹ Although the COS itself and its pronouncements on 'vice' have long disappeared (poverty is 'principally the result of a moral failure'),²⁰ the 'Sally Ann' with its 'soup runs' is still very much a mainstay of feeding the street population.

Then there has been New Labour's anxiety over street beggars using their children as props, particularly if they were asylum seekers or refugees. This was 'vile', declaimed Home Office minister Barbara Roche; 'These people have come here with the intention of exploiting the system and exploiting their children. It is a vile thing to do.'²¹ There followed an explosive debate reminiscent of the 'cadgers and imposters' variety, straight from the pages of Henry Mayhew²² or Charles Dickens. Indeed, one original motive of the COS, as of the early Poor Law, was to clear the streets of professional beggars; founded in 1869, it grew out of the earlier pressure group named the Society for the Prevention of Pauperism and Crime. Professor Mowat details COS allegations that children were exhibited by their parents to excite sympathy, having their sores deliberately induced, donning filthy rags for the day's 'work' and that begging letters became a 'profession'.²³ As with the early Poor Law moves against 'tramping', or the need for a rural police force to suppress 'tumults', or the additional powers sought in the 1930s to quell troubles at Assistance offices, so the background for New Labour is social control. As far back as the fourteenth century, there seems to have been the legislative concept of an 'army of able-bodied beggars' wandering the countryside in token pursuit of work.²⁴

Fortunately for the recipients of 'assertive outreach' under New Labour, the modern agencies involved, including the police, are not just considering the so-called 'hard options', such as the 'zero tolerance' of New York City. They have also been looking at the 'softer options' of removing individuals to wet shelters where they can get appropriate medical treatment, *boulevard* design and even all-day specialist public houses – with social and care workers in attendance.

Indeed, considerable direct expenditure is now being targeted on these 'streetlife issues', rather than having them picked up just by social security and the local NHS casualty department. The Rough Sleepers' Unit, for example, has an integrated budget of £145 million for London over three years. This funding replaces the provision of hostels, which were unified under the Resettlement Agency and then closed down. A further £34 million has been allocated for the rest of England over the same period as part of the Department of Environment's Homelessness Action Programme. To complement this, the Department of Health has a programme for grants under a Homeless Mentally Ill Initiative. In addition there have been Drug and Alcohol Specific Grants. The comprehensive spending review has now set out several funds to target the issue of 'Social Exclusion':²⁵ New Deal for

19 See for differing viewpoints, W. Booth, *In Darkest England and the Way Out* (London: Salvation Army, 1890) and B. Bosanquet, *'In Darkest England' On the Wrong Track* (1891, re-published in London: Routledge, 1906). See also R. Collier, *The General next to God: the story of William Booth and the Salvation Army* (London: Collins, 1965).

20 Sir Charles Loch, Secretary of the COS, quoted in C.L. Mowat, *The Charity Organisation Society 1869–1913* (London: Methuen, 1961) 68. The COS eventually became the Family Welfare Association, which still exists, although with rather different objectives.

21 Interview in the *Daily Mirror*, 20 March 2000.

22 H. Mayhew, *London Labour and the London Poor* (London: Griffin, Bohn, and Co. 1851, new imprint 1967).

23 n 20 above, 6. See also E.P. Thompson, *The Making of the English Working Class* (London: Gollancz, 1963) 265–6.

24 I. Jennings, *The Poor Law Code* (London: C. Knight & Co Ltd, 1936) 15.

25 Defined by the Government's Social Exclusion Unit as 'a shorthand label for what can happen when individuals or areas suffer from a combination of linked problems such as unemployment, poor skills,

Communities (£1,320 million over four years); Neighbourhood Renewal Fund (£800 million over three years); Children Fund (£450 million over three years); £1,100 million additional funds to Regional Development Agencies (over a four-year period); and doubling the number of Sure Start programmes so that they reach one-third of all poor children. Even the harshest critics of the Government have seen enormous financial commitment in this area of social policy, and although they have criticised the somewhat inchoate nature of defining 'Social Exclusion', this has been one of the most innovative areas of Government strategy.

A second key feature has been the aim of 'Welfare to Work'. Heralded as the central plank of the Clinton/Blair welfare ideology, it is of course not without its historical antecedents. Perhaps the most famous piece of British social legislation, the Poor Law Act of 1601, also had as its central theme the message of 'Welfare into Work'. The legislative demand then of 'setting the poor to work' has had many echoes in social policy, from Owenite 'Back to the Land' philosophies through to the pragmatism of Roosevelt's 'New Deal'. Traditionally perhaps, left-wing governments have had the instinct to be 'nice to the needy', but rather than handouts, Clinton's programme from 1992 pledged to 'end welfare as we know it'. This culminated in the Personal Responsibility and Work Opportunity Reconciliation Act 1996, and the plan certainly caught the attention of 'New Labour' theorists.²⁶ It has also been arguably very effective in the USA, in reducing welfare rolls by half across the nation, although there are some critics who allege that claimants have been treated harshly and that those finding work have often found it at low wage rates.²⁷ In addition, surveys conducted by the Urban Institute in Washington show that ex-welfare respondents lack health insurance, and many admitted to skipping meals or missing payments because of tight finances.²⁸

However, shifting welfare claimants into remunerative and sustained employment has proved to be a very hard task in both countries. Even at the height of the economic cycle in Britain in February 1999, nearly 6.1 million people, or 17 per cent of the working-age population, were on at least one important benefit. In some places the figure was far higher: for example, 39 per cent of those of working-age in Merthyr Tydfil, a Welsh former mining and steel town, were claimants. Under New Labour since 1997, there has been a fall in unemployed claimants, but the proportion of people receiving at least one of the main sickness and disability benefits has continued to grow. This latter group are now easily the largest single group, with 2.9 million people in February 1999.

Just as striking is the high proportion of claimants who are stuck on benefits. Over half the people who are claiming at least one key benefit were also doing so three years before. Two-thirds of these 'long-stay' claimants are drawing benefits for sickness or disability. Indeed, 77 per cent of this group of claimants, more than in any other group, have been on benefit since at least February 1996, 65 per cent of them claiming the same sort of benefits then as now.²⁹ Some might argue that

low incomes, poor housing, high crime environments, bad health and family breakdown'. The definition used by the Council of Europe is 'an inability (of individuals) to participate effectively in economic, social and political and cultural life, alienation and distance from mainstream society.'

26 See the influential work by the former General Secretary of the Fabian Society, Simon Crine, *Reforming Welfare: American Lessons* (London: Fabian Society, 1994).

27 See generally R.J. Link and A.A. Bibus with K. Lyons, *When Children Pay: US welfare reform and its implications for UK policy* (London: Child Poverty Action Group, 2000) and A. Deacon (ed) *From Welfare to Work: lessons from America* (London: Institute of Economics, 1997).

28 See for example a study of the Michigan system in the *Detroit News*, 29 December 1999.

29 See the very informative DSS, *Client Group Analysis: Quarterly Bulletin on the Population of Working Age on Key Benefits*. Just over 5.2 million (15 per cent) people of working age were claiming

sick and disabled people are bound, by their very condition, to be on benefit for long periods. But there are also disturbing overlays of health, class, gender and ethnicity issues. For example, 13 per cent of the working-age population in Wales claim a sickness or disability benefit, against 5 per cent in southeast England. Looking at London, there are vast divides; 11 per cent of the working-age population of Hackney draw a sickness or disability benefit, compared with 5 per cent in the Royal Borough of Kensington and Chelsea. These huge disparities, particularly in health, have an enormous impact on welfare expenditure.³⁰ They were, of course, a feature discovered in the pioneering study of London Poverty by Charles Booth in 1889,³¹ but dealing with this multi-faceted problem is no easier now than then. Interestingly, sickness absence rates and self-reported disability show regional variations, and are not just based on class and income disparities. There therefore seems to be a cultural propensity to regard oneself as disabled, acceptable in some communities and stigmatised in others. However, some commentators have suggested that '[t]he truth is that some of Britain's vaunted fall in unemployment has been achieved by shuffling people into benefit categories where they have less incentive to work, and are less likely to be chased by officialdom to look for jobs.'³² Hence New Labour's (perhaps reluctant) drive to pursue the 'Work to Welfare' mantra amongst those who are chronically sick and disabled, where it becomes very much more problematic and open to criticism, rather than just among the alleged 'workshy' in the ranks of the able-bodied unemployed. *Labour Force Survey* figures suggest that over half of the 2.6 million economically inactive people on Incapacity Benefit say they want to work. Saying this, and being exposed to the potential for disqualification, may not be entirely compatible. The latest proposal from the Chancellor, to compel married housewives under 45 with no children to take part in the New Deal jobs programme is similarly being approached with caution. This compulsion would affect an estimated 55,000 women, but will not come into effect until 2002. Currently only 43 per cent of partners of unemployed men are in work. Somewhat predictably, family and marriage groups immediately accused the Government of 'hounding' women out of the home against their will.³³

A third central theme of welfare systems has been the desire to 'control fraud'. David Miliband makes the point that 'Any Government that wants to modernise the welfare state to meet new needs has to convince taxpayers that it is affordable ... it is why action against fraud has to be taken forward across government.'³⁴ Again, little seems to have changed in the debate since the 'eternal principles' of

a key benefit in February 2000. This compares with 5.9 million in February 1997. Those claiming sickness and/or disability benefits are now by far the largest group (2.9 million in February 2000).

30 See the report by the Association of London Government, *Sick of being excluded* (2000), examining the literature on why black and ethnic minority communities carry a higher burden of poor health, premature deaths and chronic long-term ill health and disabilities. See also Directorate of Public Health for East London & The City Health Authority, *Health of Ethnic Minority Elders in London* (2000).

31 C. Booth, *Labour and Life of the People of London* (17 volumes, Williams and Norgate, 1889–1903). Booth's maps of 1889 show somewhat quaint classifications ranging from the yellow of 'Upper-middle and Upper classes. Wealthy' down to the black of 'Lowest class. Vicious, semi-criminal', but the geography bears an uncomfortable resemblance to the neighbourhoods targeted in the Government's Regeneration Strategy.

32 *The Economist*, 1 December 1999 3.

33 *The Times*, 1 December 2000. 'The Government is not interested in providing choice but in fulfilling their own policies', said Valerie Riches, president of Family and Youth Concern. 'Men are being pushed to the margins of society. Their masculine role as the provider and father is being obliterated.'

34 Miliband, n 16 above.

1834, when a previous reformist government had been swept into office and set about clearing up abuses endemic in the Old Poor Law. The Benthamite, or perhaps rather Chadwickian, solution adopted by the Whig Government then led to incalculable cruelty in the workhouse system, in a futile attempt to 'discourage the poor from their poverty'.³⁵ The social disease diagnosed by the Royal Commission was 'pauperism', or an unwarranted dependence on the state. When the former New Labour Minister for Welfare Reform selects as the title of his book, *The State of Dependency*, then it is clear the debate is still moving along familiar lines, even if the kaleidoscope has been shaken.

Fraud is clearly present at some level in any system of public finance. However, since the graphic descriptions of the 1834 Report, which uncovered innumerable instances of abuse in the workhouses, the payment of 'out relief' and in the 'Roundsmen' system,³⁶ mythology has perhaps exaggerated the true position. Corruption then was endemic. Now perhaps it is more individualistic. For example, one of the most extraordinary fraud cases of recent times was that of David Hill in 1999. Known to the DSS as a housebound invalid requiring 24-hour care from social services, and £43,000 in benefits for his crippling asthmatic condition, Mr Hill appeared also to be an armed robber, who engaged in scuba diving, sailing, skiing, running half marathons and, as the holder of a pilot's licence, flying to exotic locations to pursue his hobbies. Mr Hill is now serving a 20-year prison sentence for conspiracy to commit robbery and defrauding the DSS.³⁷ His widely reported case inevitably fuelled myths about the extent of abuse.

The official estimate of social security fraud for the past three years has settled on the estimate of £4 billion per annum, or 'enough to give every man, woman and child in the country an £80 handout' according to a report by the Public Accounts Committee of the House of Commons. They stated that their fraud estimate was 'unacceptably high' and expressed frustration at the slow progress of government anti-fraud measures. Shortly before publication in August 2000, the DSS announced that they had drafted in John Alpass, Whitehall's intelligence co-ordinator and former deputy director-general of MI5, to take charge of intelligence gathering in a new anti-fraud branch. Mr Alpass is due to work closely with police and undercover agents to monitor suspected fraudsters, and will send in 'hit squads' to help struggling anti-fraud units in local benefits offices.³⁸ The ubiquitous Mr Field has also stepped into this debate by focusing on housing benefit, claiming that the 'refusal' of some local authorities has left many 'providing safe havens for fraudsters'.³⁹

As well as the philosophical issues, political considerations are also important; all polling evidence shows considerable support for any government's standing firm against 'scroungers' and 'malingerers' on behalf of the taxpayer. It was therefore fairly unsurprising that in the December 2000 Queen's Speech there was confirmation of a Social Security Fraud Bill, with powers to force banks to pass on details of suspected fraudsters, and an extension of powers to the Inland Revenue and Social Security fraud investigators.⁴⁰

35 R. Mendelsohn, *Social Security in the British Commonwealth* (London: Athlone Press, 1954) 23.

36 See S. and B. Webb, *English Poor Law History* (London: Longmans, Green and Co, 1927) vol 1, 190.

37 *The Times*, 1 December 1999.

38 *The Times*, 9 August 2000. The initiative will co-ordinate the work of 5,000 staff, including 3,700 investigators.

39 *The Times*, 11 April 2000.

40 *The Times*, 7 December 2000.

Contextual tensions

Welfare legislation does not exist in a vacuum, and there are many important organisational issues. Again, there are some well-travelled roads. The first issue is financial, but not so much the global figures as the recoupment between different areas. In particular, there is a complex interface between social security and health, which are national budgets, and then the inter-relationship of both agencies with local authority social services. This has been a never-ending argument, as budget holders fight to retain control of their own domains.

Take for example the everyday event of a retirement pensioner suffering a fall. Approximately half the women in Britain will suffer a fracture in this way at some time in their lives, and hospital admission is obviously the way to deal with an acute problem. Health Service performance indicators will look for the earliest possible exit from a much-needed and expensive hospital bed. For someone over retirement age, the quickest route out will be to an Elderly Persons' Home, run by a local authority or financed by them in the independent sector. But, as is now clear, this 'dumping ground' may be wholly inappropriate in the long term. If budget considerations in the Health Service are not to take priority over civil liberties and long-term stability, then a careful and planned discharge, perhaps with an intermediate period of respite care or convalescence, will see the pensioner re-established in their own home, perhaps with support. If the claimant has a private pension or private health insurance, then this longer road may be possible. And if the pensioner lives in a local authority with serious home care services, or has a devoted family (occasionally assisted themselves with respite care), then the pensioner may also take the optimal route back to being self-sufficient. These micro-decisions on personal welfare have enormous impact on budgets.⁴¹

The Poor Law had its own labyrinthine sets of recoupment problems, between parishes and Unions intent on reducing expenditure for their ratepayers. Indeed, in 1909 there were still over 10,000 cases annually on the Laws of Settlement and Removal.⁴² These draconian laws determining the residence of paupers clogged the courts. Disputes on who should pay 'out relief' disappeared on unification of the system, following the establishment of the Unemployment Assistance Board in 1934 for the country as a whole. However, in 1909 the Poor Law included hospitals and schools, so as the 'Break Up' occurred of this monolithic organisation, the budgetary boundary disputes reared up in other areas. In the modern era, the interface between health and social services is still a fertile field for 'brokerage' and negotiating skill, but the Government have also had to deal with costs arising from refugees, asylum seekers and economic migrants. These groups, like many others in the welfare field, are very difficult to compartmentalise, and lobbying on the resultant legal and care costs has been intense.

This leads on to a tension between atomising services for discrete groups, or attempting 'joined-up' government. This continuum has its fashions. The Webbs in their 1909 Minority Report urged 'Break Up' because of the incapability of the Poor Law to deal with a vast array of social problems.⁴³ The trend in more recent years has been to attempt better co-ordination in Whitehall, and at least more 'joined-up' thinking between departments, but this has been notoriously difficult to

41 See generally H. Glennerster and J. Hills (eds), *The State of Welfare* (Oxford: OUP, 2nd edn, 1998).

42 See generally chapter five in P. O'Connor, *Vagrancy* (London: Harmondsworth, 1963).

43 S. and B. Webb, *The Break-Up of the Poor Law* (London: Longmans, Green and Co, 1909). See also M. Cole, *The Webbs and Their Work* (London: Muller, 1949).

achieve. The Labour Government has now had several 'enforcers' (David Clark, Peter Mandelson, Charles Falconer, John Cunningham, and Mo Mowlam), which perhaps illustrates how difficult it is to achieve this at the national level. Similarly at the local level, despite the endless meetings and case conferences, there continue to be many financial questions and huge disparities in service, focused on the current mantra of 'Best Value'. The *Quality Strategy for Social Care* states that 'Some of the most serious complaints by users of social services are about inconsistencies and variations in services between one part of the country and another – often between neighbouring councils.'⁴⁴

A classic example of the difficulties on 'joined-up' thinking is the policy on rough sleeping. Section 30 of the Jobseekers' Act 1995 enables discretionary grants to be made to providers of temporary supported accommodation with an active resettlement programme for single homeless people. These are mainly voluntary organisations, such as St. Mungo's, but also registered social landlords and nine local authorities. A prime mover in policy on these issues is the government's Social Exclusion Unit. They produced a report on rough sleeping in July 1999 which made recommendations, perhaps predictably, for a 'more co-ordinated approach toward policy and operations' on rough sleeping across Whitehall. There were clearly many interests involved here: the health services, local authority housing and social services departments, highway authorities, parks and leisure departments, the voluntary sector providing shelter and outreach, housing associations, social security, and so on. At the national level, a decision was taken to shift responsibility for resettlement funding in England from the Department of Social Security to the Department of Environment from April 2000, except in London, where it had already been transferred on the setting up of the Rough Sleepers Unit. The resettlement funding programme, with a total budget over £18 million for 1999–2000, now supports over 4,200 bed spaces provided by over 100 organisations managing some 260 projects for single homeless people in England, Scotland and Wales. And this is just one area looking at Social Exclusion amongst the street population, with other initiatives on street drinkers, beggars, drug users and sex workers.

The third related issue is the vexed partnership between the public and the private sectors. Again, this is well-charted terrain in welfare provision. On so many occasions, a British social institution has been set up as a charitable experiment, and then has become part of the public infrastructure of services. The list includes the first Workhouse, set up in Bristol in 1697; almshouses which have slowly changed into Housing Associations, funded by the Housing Corporation; social workers who transmogrified from the almoners and outreach workers of the Charity Organisation Society; probation officers and prison visitors, from the work of concerned citizens and members of the Society of Friends; Children's Homes and adoption services, from the work of Captain Coram and Thomas Barnardo. Beveridge himself, after his famous report in 1942, went on to consider the partnership of state and charity in *Voluntary Action*,⁴⁵ but the complex interplay of public and private is even more a feature of public debate, not least because the Conservative Party have now announced their intention to 'privatise' industrial injury benefits on a return to government.

⁴⁴ Department of Health, *A Quality Strategy for Social Care* (2000) 3.

⁴⁵ W. Beveridge, *Social Insurance and Allied Services* (1942), *The Pillars of Security* (1943), *Full Employment in a Free Society* (1944), *Voluntary Action* (1948).

The legislation

As with many statutes it will take time to assess the effectiveness of the 1999 and 2000 statutes. In social security law there is also the long shadow of delegated legislation which can be very potent. The key words in many such Acts are the springboard phrases 'The Secretary of State may by regulations provide that ...'. For example, in looking at the way Incapacity Benefit claimants are 'actively' encouraged to return to work, the decisive arena is not so much the legislation of either of the last two governments, but the Social Security (Welfare to Work) Regulations.⁴⁶

The framework of the Welfare Reform and Pensions Act 1999 deals with some relatively non-controversial issues, such as the new stakeholder pension schemes; pension sharing for divorcing couples; and updating benefits for widows and widowers. With the uncertainty of predicting Human Rights Act litigation, perhaps nothing is completely guaranteed to be 'non-controversial'. Everything is relative. But as we have seen, the 'modernising' of benefits for those with disabilities or long-term illness caused acute difficulties for the Government, not least because it unleashed Labour peers such as Jack Ashley and Barbara Castle. But probably the most important change in the Act has been the introduction of the 'single gateway' to the benefit system. This was piloted successfully in several parts of the country. The aim is to respond to research showing that many claimants are mystified in trooping from one office to another. In response there will be a work-focused interview as a condition of benefit entitlement.⁴⁷ Waivers will be a key issue, as are 'disregards' generally in the social security system. The quality and training of the officers who carry out the interviews will also be vital. The Government's aim is to provide claimants with a 'single point of contact' and if this works across the divide with, for example, Council Tax and Housing Benefit, then it will be very productive. 'Personal Advisers' are there to aid the mentoring process, and innovative mechanisms with a 'call centres model' and a 'private/voluntary sector model' have been introduced. If this works, it will be on a par with the Beveridge/Churchill model of employment exchanges, which was such a breakthrough in 1910.

The second statute, the Child Support, Pensions and Social Security Act 2000, has a focus on children. Again, this chimes in with repeated pronouncements by Government ministers. Backing them at the No. 10 Policy Unit, David Miliband confirms that 'Children are a group for whom a top class welfare state is essential, but who have been let down ... Check any benchmarks for welfare policy – investment, social mobility, tackling poverty – and the needs of children are a neat fit.'⁴⁸ The Act therefore takes on the concerns about the Child Support Agency, set up to deal with the old 'liable relative' rules in social security. Scarcely a day goes by without a newspaper story on suicidal fathers harried by the CSA, and there have been many criticisms of the organisation's mathematics and unrealistic demands. The CSA currently handles maintenance for more than 970,000 families and in 1998/99 arranged £661m of child maintenance. The Prime Minister has also made it clear that 'We cannot have taxpayers as a whole ending up having to pay large sums of

46 SI 1998 No. 2231. See further n 13 above.

47 See D. Allinajah, 'The Single Work-Focused Gateway, aka the 'ONE'. how it works' *Welfare Rights Bulletin*, issue 150, June 1999, 4.

48 Miliband, n 16 above, 12.

money out to help families in circumstances where their fathers don't have any sense of responsibility.'⁴⁹

The Act gets tough with recalcitrant fathers, with its 'one strike and you are out' policy, under which absent fathers who miss even a single payment will have the owed money deducted directly from their salary. But it also imposes a tariff arrangement of 15 per cent of net income for a first child, 20 per cent for two, and 25 per cent for three or more. This will give firmer rules known to all the players, and will hopefully reduce the vast expenditure of effort on calculations by the CSA.

The Act contains several examples of lateral thinking, with a novel crossover between the workings of the welfare and the criminal justice systems. Offenders who fail to comply with community sentences passed by the courts will be liable to have Income Support, Jobseeker's Allowance, and Training Allowances reduced or temporarily withdrawn. Income Support will be reduced by up to 40 per cent for four weeks and the Jobseeker's Allowance can be withdrawn altogether, after one written warning. Most interestingly, under section 16 there are powers to take away the driving licences of fathers who fail to pay maintenance. A court determination on breach will be needed before these powers are invoked, and the whole scheme is subject to being piloted first. There will also be powers to ensure that assessments cannot be delayed unnecessarily by disputes over paternity, allowing presumptions of parentage where an absent parent refuses a DNA test, or refuses to accept the positive outcome of a test.

On pensions, the Act takes forward the updating of pensions, encouraging employers to provide more user-friendly schemes, and introducing the State Second Pension to help low and moderate earners save for their retirement. This reform of the State Earnings-Related Pension Scheme (SERPS) continues work started in the Crossman era of the 1960s. The Act also gives pension rights to carers and disabled people with broken work records. The Government estimates that around 11 million people will be better off when they retire as a result.

However, the new Act has not been without its critics. The plan to dock the benefits of the so-called community sentence 'no-shows' was repeatedly condemned by the National Association for the Care and Resettlement of Offenders, who called it 'retrograde and counterproductive.'⁵⁰ Professor Steve Webb MP, the formidable Liberal Democrats' social security spokesman, also dismissed the device, claiming that 'Cutting off the only source of income from offenders is very likely to drive them further into a life of crime. This does nothing for the victims of crime or to stop re-offending.'⁵¹ In response to these views, the approach will be piloted in various parts of England and Wales to test the links between social security offices and the Probation Service and to assess the behavioural impact on offenders. Professor Webb also criticised Labour's plans for reform of the CSA, arguing that the agency should be abolished.

49 *The Scotsman*, 18 November 1999. Alistair Darling claimed that about 20 per cent of the annual caseload of 125,000 offenders in England and Wales failed to comply with conditions: *Daily Telegraph*, 2 December 1999.

50 *The Herald*, 2 December 1999. NACRO's Director, Paul Cavadino, said: 'It makes no sense as a way of tackling crime. Plunging offenders further into poverty must increase the temptation to commit theft, burglary or street robbery, so damaging, not improving, community safety.' The National Association of Probation Officers also expressed concern; 'It is extremely difficult to envisage the withdrawal of benefits from those who breach orders either enhancing public protection or reducing crime.'

51 *The Scotsman*, 18 November 1999.

The next phase

The Government's intention is to press on with welfare reform, so another round of measures was a possibility for this year's Queen's Speech. However, given the pressing claims of other political demands, the Government has opted for the low-key Social Security Fraud Bill, which also sends out reassuring messages in the run-up to the General Election. But what are the options in the longer term? David Walker has suggested that 'Welfare reform is a permanent condition: the system constantly needs updating to cope with changes in family life let alone the labour market.'⁵²

One intricate matter is Housing Benefit, long a target area for reform. The Beveridge Report pointed out the critical distinction of housing costs for claimants, and the old housing allowance in National Assistance and Supplementary Benefits tried to deal with disparities. However, the response of the Thatcher Government was to offload this responsibility to local authorities, where the experience has not been entirely happy, and where there are grounds for believing that the system has been milked by serious criminal effort. The cost of housing benefit has doubled since 1990 and trebled since 1986, partly because of large rent increases due to the last government's decision to deregulate rents. As rents are paid virtually in full by local housing departments, who then claim back the cash from the Department of Social Security, tenants have no obvious incentive to seek cheaper accommodation. However, reform of the benefit would also mean intervention into the housing market, which may trigger many undesirable consequences for local and central government; it would be a long-term strategy and would possibly not save money.⁵³ Housing Benefit costs £11 billion a year, accounting for almost one-eighth of the entire social security budget, so it is an attractive target, and the recent Green Paper on Housing suggests a 'streamlining'.⁵⁴ Some 40 per cent of claimants here are retirement pensioners, and 60 per cent live in council accommodation. The focus is likely to be on housing benefit fraud, estimated to be running at up to £600 million a year, by simplifying the system and centralising control through nationwide computerisation. The Prime Minister's characterisation of Housing Benefit as one of the most 'complex, unpredictable and inconsistently administered' parts of the welfare state⁵⁵ certainly indicated an intention to tackle the area if at all possible. The Conservatives have made this a focus of debate; suggestions have been made that the Government should give a flat-rate housing benefit directly to all low-paid people, which it is claimed, would cut fraud and save the Government up to £350 million a year.⁵⁶

Another major project would be the further dismantling of the contributory/non-contributory divide in social security. A feature of the system since the Beveridge National Insurance Act in 1911, further intensified in his work as the Chairman of the Unemployment Insurance Statutory Committee in the 'Double Decker' system after 1934,⁵⁷ and then reified following his 1942 report, the division looks

⁵² Walker, n 6 above.

⁵³ See further *Housing Benefit: time for reform* (Joseph Rowntree Foundation, July 1998).

⁵⁴ *Quality and Choice: A Decent Home for All* (April 2000), describing itself as 'the most comprehensive review of housing in England for nearly a quarter of a century'

⁵⁵ *The Independent*, 6 December 1999.

⁵⁶ P. King, *Reforming Housing Benefit. What the Government Ought to Do, But Won't* (The Adam Smith Institute, 2000). Dr King claims that direct payment would prevent unscrupulous landlords claiming for non-existent tenants and save millions in administrative costs and payment errors. Latest estimates show housing benefit expected to increase to £12.4 billion by next year with more than half of all tenants claiming.

⁵⁷ See W. Beveridge, *The Unemployment Insurance Statutory Committee* (1937).

increasingly unrealistic. To be fair, Beveridge had the notion that the Assistance side of social security would 'wither away' and there was a wide gulf between the principles of the Beveridge Report and the Attlee Government's implementation of the scheme. Call it 'means-tested' or 'needs-based', the whole thrust of New Labour has been away from universalism and towards targeting benefits. In response to a critique from Professor Ruth Lister, Alistair Darling makes clear that in his view

We must concentrate effort to those who need the most help: three million people workless for more than two years; the 1 in 5 children living in households where no one works; the many children who leave school with no qualifications; pensioners isolated within their own communities. Today the important difference in social security systems is not whether they are insurance based or means tested, but whether or not they provide enough help to get to people back to work and improve their lives.⁵⁸

Dismantling the Beveridge contributions and benefits edifice in one swoop, still surprisingly just about in place, even though parts of it have crumbled, is probably not an option. There would be too many ramifications. 'Radical reform of the welfare state' has again been promised in *The Changing Welfare State: Social Security Spending*, a DSS document issued in February 2000, but the main emphasis was to insist that the social security budget was under control, with a prediction of growth by just over 1 per cent above inflation, assuming macro-economic stability.⁵⁹ However, nibbling further at aspects of universalism, perhaps even the sacred area of Eleanor Rathbone's Family Allowances, which evolved into Child Benefits, may be a possibility. Some commentators feel that this is underway with child tax credits. In terms of the total social security budget Child Benefit is not huge: about 8 per cent, which is about the same as for Job Seeker's Allowance, compared with 26 per cent on disability and 46 per cent on pensioners. It is the latter two groups where numbers have been soaring, with a strong demographic push for those above retirement age. Political considerations here about a backbench backlash may deter a direct assault, but unattributable opprobrium has already been heaped on this 'middle-class' benefit.

Probably the area most under Government scrutiny is that of industrial injury benefits. Often seen as the pacesetter for other benefits, it was described by Tony Lynes as the 'sacred cow' of social security. With a history going back to the Workmen's Compensation Act of 1897, the industrial injury benefits system was created in 1946 following the Beveridge Report. Subjected to sustained battering in the Fowler Reforms of the previous Conservative government, it has been the subject of various proposals to privatise or modernise the structure. The suggestion has been that this is an 'anomaly' set up to deal with employees in nationalised industries, and should now more readily be dealt with as part of the tort system, where the underpinning reality is insurance.⁶⁰ Disablement benefit is paid to 282,000 people and costs £712 million every year, so potential savings are relatively modest. The Department of Social Security held informal discussions with the CBI, the TUC and the Association of British Insurers through 1999, but fierce opposition to any extra burdens on business has been revealed. However, David Willetts MP, the shadow social security secretary, has already proposed the privatisation of industrial injuries payments, terming the present system 'outdated

⁵⁸ *The Guardian*, 16 June 1999.

⁵⁹ See analysis in *The Independent*, 10 February 2000.

⁶⁰ See generally chapter 13 in P. Cane, *Atiyah's Accidents, Compensation and the Law* (London: Butterworths, 6th ed 1999).

and pernicious', on the basis that there is no incentive for employers to improve their health and safety record.⁶¹ Probably with enough on the current agenda, this may be a matter left on a back burner until after the next election, when union financial support remains vital; the TUC itself is not opposed to reform of the system, but they are of course adamantly against cuts in benefits.

Conclusion

As with the terms 'Old Poor Law' and 'New Poor Law', it is possible to point to specific legislative changes carried out by New Labour in the welfare field. But the continuities in the kaleidoscope are overwhelming. Several of the 'new' ideas, such as urging the public not to give charitable donations to beggars, but to channel support through organised channels, are breathtakingly familiar. As ever, the philosophical stance underpinning such a policy is hotly contested. Indeed, perhaps the only 'new' part of such a policy is that the evil allegedly supported by such indiscriminate charity is the menace of drugs rather than drink.

Looking at the governmental programme in terms of periodicity, the 'modernising' approach started out with the brainstorming first year of 'thinking the unthinkable'. It ended with a visionary list of 32 'success measures' but with the leap overboard by the key Minister for Welfare Reform. Mr Field's trajectory through this process has been instructive, because from being chief architect he has moved to guerrilla activist, sniping from the sidelines both on the content and the efficacy of the reforms. The first welfare bill of 1999 was the battleground, but with Alistair Darling's subtlety and advocacy skills, the Government's programme survived largely intact. The second statute in 2000 was therefore tempered by the knowledge of the sort of political firestorm that might be encountered. However, its measures on child support and pensions are innovative, and may indeed confirm the prediction of being lasting 'milestones' in this welfare reform process.

Nevertheless, the key themes on dealing with rough sleepers, beggars and the control of fraud are long-term and intractable matters that even New Labour is finding difficult. The Rough Sleepers' Initiative has certainly assisted an important and often forgotten handful; the 'cardboard city dwellers' which were a feature of the previous Thatcher/Major years are now largely indoors. But the legislation, the edicts and the policy of New Labour to pursue 'the street population' has an oddly authoritarian edge.

This is perhaps even more pronounced in the Government's attitude to fraud, where myth is very difficult to separate from reality. Consciously needing to avoid being outflanked by the Conservative opposition, and needing to appease popular opinion, the Government is determined to show its mettle. Needless to say, the tabloid press are always very eager to supply examples of allegedly outrageous behaviour.⁶² The reality is that fraud is likely to be present in any public benefit scheme, but the inordinate amount of time spent on this initiative has not yet shown results. Critics chide the Government for its lack of resolve, but this is unlikely. The principal difficulty is the scale and complexity of the issue. Possibly new technology, just as with preventing the counterfeiting of money, can assist. But this is a matter of eternal vigilance, and New Labour is likely to remain in the usual governmental 'no win' position.

61 *Daily Telegraph*, 24 December 1999.

62 See for a classic of the genre, K. Weitz, 'She's got 11 kids and hasn't had a job in 20 years ... yet this woman is raking in a massive £33,334 every year milking the system' *The People*, 15 October 2000.

One interesting aspect of the New Labour Government is the Scottish dimension. As well as the Prime Minister's education and ancestry in Scotland, there are very significant Scots in the Cabinet: Gordon Brown, Lord Irvine of Lairg, Robin Cook, and of course Alistair Darling. The legacies of a former leader, John Smith, and a former Chief Whip, Donald Dewar, also linger from welfare policy formulation in Opposition. The different trajectory of the Scottish Poor Law,⁶³ as well as different cultural assumptions, may be seen to have long-term influence.

New Labour has certainly not lost its impetus or appetite for a 'modernising' agenda. The current period is inevitably a pre-election phase, when legislative rather than fiscal prudence is advisable. Housing Benefit, the contribution principle, Industrial Injury Benefits, have all been in the spotlight, and the Government if re-elected will no doubt return to these issues in future legislation. But the practicalities of a large legislative programme and a run up to a General Election probably dictate a pause while the Acts of 1999 and 2000 are digested and implemented. What seems inevitable, with this or any other government, is a further wave of welfare legislation.

63 See R.A. Cage, *The Scottish Poor Law 1745-1845* (Edinburgh: Scottish Academic Press, 1981).

REPORTS

New Labour, Same Old Tory Housing Policy?

Dave Cowan and Alex Marsh***

For the first time in 23 years, the government has produced a Green Paper on housing – *Quality and Choice: A Decent Home for All* – which has ‘set out [the New Labour] vision for housing in the new Millennium’.¹ In the intervening period, housing policy has not remained static – indeed, the landscape of housing policy has been dramatically altered by a series of White Papers.² The role of local authorities has shifted towards ‘enabling’, rather than providing, new social housing; housing associations – which, along with local housing companies, are now termed Registered Social Landlords (RSLs) – have become major providers of social housing; the private rented sector has been deregulated; considerable levels of private finance have been levered into the social housing system; demand for social housing has bottomed out in certain areas; mortgage finance of owner occupation has undergone dramatic shifts as a result of deregulation during the 1980s.³ New Labour itself, prior to the Green Paper, has progressed its own housing-related policies such as Best Value,⁴ alteration of the local authority financial regime,⁵ allowing councils to phase in the spending of capital receipts from the sale of their stock on specified projects,⁶ and various initiatives derived from the Social Exclusion Unit concerned with neighbourhood renewal.⁷ Indeed, the government is keen to locate the Green Paper as one policy statement in an ongoing stream of policy developments.⁸

* Department of Law, University of Bristol.

** School for Policy Studies, University of Bristol.

1 (London: Department of Environment, Transport and the Regions [DETR] 2000) para 1.3; its predecessor was Department of Environment (DoE), *Housing Policy – A Consultative Document* Cmnd 6851 (London: HMSO, 1977).

2 DoE, *Housing: The Government's Proposals* Cm 214 (London: HMSO, 1987); *New Financial Regime for Local Authority Housing in England and Wales: A Consultation Paper* (London: DoE, 1988); *Our Future Homes: Opportunity, Choice and Responsibility* Cm 2901 (London: HMSO, 1995).

3 Compare the concerns in D. Cowan, *Housing Law and Policy* (Basingstoke: MacMillan, 1999) with those in A. Murie, P. Niner and C. Watson, *Housing Policy and the Housing System* (London: Allen & Unwin, 1976).

4 Local Government Act 1999; as to which see P. Vincent-Jones, ‘Central-Local Relations under the Local Government Act 1999: A New Consensus?’ (2000) 63 MLR 84.

5 *Modernising Local Government: Capital Finance*, Consultation Paper (London: DETR, 1998); *A New Financial Framework for Local Authority Housing: Resource Accounting in the Housing Revenue Account* (London: DETR, 1998).

6 *Capital Receipts Initiative: Guidance to Local Authorities*, (London: DETR, 1997).

7 *A National Strategy for Neighbourhood Renewal* (London: Cabinet Office, 2000); see also the various reports of the Policy Action Teams especially the following: *Neighbourhood Management* (London: SEU, 2000), *Anti-Social Behaviour* (London: SEU, 2000); *Report by the Unpopular Housing Action Team* (London: SEU, 1999).

8 See, for example, the subsequent DETR, *Our Towns and Cities: The Future – Delivering an Urban Renaissance* (London: DETR, 2000).

The last Conservative government's housing White Paper was subtitled 'opportunity, choice and responsibility'. Opportunity concerned individual and private sector entrepreneurialism, particularly through new technologies; individual choice was to have been maximised through housing policy; and responsibility concerned empowering individuals to make their own choices rather than rely on the state.⁹ These key themes for housing policy are equally evident in the New Labour Green Paper, both in terms of their content and underlying message. So, for example, the 'key principles' of the New Labour housing policy are said to be, *inter alia*, 'offering everyone opportunity, choice and a stake in their home, whether rented or owned'; and 'giving responsibility to individuals to provide for their own home where they can, providing help for those who cannot'.¹⁰ Thus, despite the rhetoric, there is historical continuity in the themes of housing policy. As we argue below, there is also continuity in much of the substance of housing policy.

The Green Paper probably comes at an important juncture for housing policy. Since the 1995 White Paper, housing academics have debated whether housing policy even has a future.¹¹ The conclusion seems to be that, if it does have a future, it will be very different from its past. This is because a number of dramatic shifts during the twentieth century have combined to alter the conception of housing policy: the retreat from public sector provision of housing, increasing reliance on private finance and markets, global economic structures, shifts in the consumption of housing towards ownership and away from renting, and shifts in finance away from 'bricks and mortar' towards individual subsidy.¹² Housing policy itself is subservient to other concerns generated by the labour market, social security, social exclusion, and economic policy and it is within these policy areas that much policy relevant to housing is now formulated.¹³

This contingent nature of housing policy means that solutions are not readily available through housing policy itself. The Green Paper reflects this through its 'greenness' – that is, policy proposals are simply not formulated in many crucial places. On the crucial questions of the day – what to do about housing benefit and the poverty trap? what to do about rents in the social sector? what to do about standards in the private rented sector and for owner-occupation? – concrete proposals are relatively modest or a range of options are advanced for discussion and debate. Where solutions are proffered, generally these are extensions of policy directions set by the previous government (such as the proposed sale of council stock through stock transfer,¹⁴ 'local lettings',¹⁵ strategic role of local authorities,¹⁶

9 *Our Future Homes: Opportunity, Choice and Responsibility*, Cm 2901 (London: HMSO, 1995) 3.

10 n 1 above, para 1.5.

11 G. Bramley, 'Housing Policy: A Case of Terminal Decline?' (1997) 25 *Policy and Politics* 387; P. Malpass, 'The Unravelling of Housing Policy in Britain' (1996) 11 *Housing Studies* 459; P. Malpass, 'Housing Policy: Does it Have a Future?' (1999) 27 *Policy and Politics* 217, and subsequent critiques by M. Kleinman, 229 and P. Williams, 231. The White Paper was regarded as a 'revealing document for it shows quite clearly the lack of policies designed to address current problems in the housing market': Malpass (1996) 468.

12 Perhaps the most dramatic example of this shift came in the early 1990s after the failed attempt to shadow the Deutschmark caused interest rates to rise and subsequent mortgage failure to occur. For an overview, see D. MacLennan and G. Pryce, 'Global Economic Change, Labour Market Adjustment and the Challenges for European Housing Policies' (1996) 33 *Urban Studies* 1849.

13 Bramley, for example, refers to 'a whole range of *de facto* housing policies [such as social security, planning, financial regulation, private finance, regeneration] masquerading as something else': n 11 above, 403.

14 A policy which has been ongoing since 1986.

15 See Housing Corporation, *Performance Standards* (London: Housing Corporation, 1997) Standard H2.2.

16 See DoE, *Housing: The Government's Proposals*, Cm 214 (HMSO, 1987) ch 5

and voluntary regulation in the private sector).¹⁷ There is a sense of powerlessness at the heart of housing policy, partly, as we argue below, because of the received set of wisdoms through which housing policy is framed.

In the first section below, we provide an overview of the proposals contained in the Green Paper. In no sense can this be complete – there are a number of small technical changes suggested which cannot be dealt with here;¹⁸ our intention is to show how the themes of quality and choice are represented together with their substantive implications. In the subsequent section, we discuss certain specific proposals in the Green Paper against a number of themes of housing and more general social policy concerns: the positioning of the state in housing provision; central-local relations; rules and discretion; rights and responsibilities; and the use of strategies of control and persuasion. In the third section, we briefly look at some of the issues with which the Green Paper does not engage and explore the reasons why that might be.

The Green Paper – an overview¹⁹

Quality and choice are the key themes of the Green Paper, but they do not exhaust its scope. The concern with quality encompasses all tenures and is directed variously at the standard of existing and new build stock, housing management, construction and building industries, and financial products. The Green Paper's discussion of choice is more focused upon the social rented sector and, in particular, upon the issue of allocations and transfers. For an extended period there has been substantial underinvestment in the housing stock across the major tenures. This has led to problems of disrepair and the need for extensive improvement and modernisation. A key current policy question is how to ensure that the standard of the existing stock is raised. The previous government took the view that the maintenance of standards in the private sector is largely a matter for individual owners and landlords. Yet, it is clear that some individual property owners either cannot, or choose not to, invest adequately in their properties.²⁰ In the council sector, the Conservative government presided over the emergence of a major backlog of repairs by restricting borrowing permissions and the release of capital receipts. This in turn led many local authorities to view the transfer of their stock to the RSL sector as the only route to substantial re-investment.²¹

An early move by the current government was to make available £800 million in capital receipts for re-investment in the council stock over the period 1997–99.²² This was followed by an increase in planned spending for housing for the period

17 See, for example, DETR, *Cowboy Builders: A Consultation Paper* (London: DETR, 1997); Council of Mortgage Lenders, *Code of Practice on Mortgage Lending* (London: CML, 1998).

18 For example, technical changes are suggested to the homelessness legislation including extending the categories of 'vulnerable' persons for priority need to include those with an institutionalised background and those fleeing domestic violence (para 9.55) – changes which should be generally welcomed.

19 The following is a summary of the main proposals. There are a number of other proposals which the Green Paper touches upon but which are not considered here (for example, leasehold reform – paras 4.17–4.19 – policy towards low-cost owner-occupiers – paras 4.32–4.41 – and possible changes to benefit help with mortgage interest payments – paras 4.41–4.55).

20 See P. Leather, *Crumbling Castles? Helping Owners to Repair and Maintain their Homes* (York: Joseph Rowntree Foundation, 2000).

21 D. Mullins, P. Niner, and M. Rieborough, 'Large scale voluntary transfers', in P. Malpass and R. Means (eds), *Implementing Housing Policy* (Buckingham: Open UP, 1993).

22 See DETR, *Capital Receipts Initiative: Guidance to Local Authorities* (London: DETR, 1997).

1999–2002 by £3.9 billion.²³ While this injection of funds has been welcomed we should not anticipate that public spending will resolve the issue of disrepair in the sector. The Green Paper envisages that, in order to generate a sufficient increase in the quality of the stock, alternative routes will need to be pursued. Specifically, the government sees a major role for the continuation of the programme of transferring stock from the local authority to the RSL sector that began under the Conservative government. The potential of the Private Finance Initiative (PFI) as a vehicle for bringing investment into social housing is also noted, although at present this is largely an area for speculation because the PFI is much less well developed in housing than in areas such as health and transport. The third route for increasing investment is the proposal for the creation of a new type of organisational arrangement – the local authority-owned arms-length management company.

The Green Paper sets out the merits of stock transfer and an ambitious set of proposals for continuing and extending the transfer programme.²⁴ If its vision is realised then, by 2004, RSLs would become the majority providers of social housing. A range of reforms directed at improving the transfer process are also presented. The option of forming an arms-length management company will allow local authorities to 'retain and use more of their rental income to finance borrowing for investment in stock improvements'.²⁵ This option will not be open to all but to the best performing local authorities that have demonstrated their efficiency in management. Even then the process will, like the stock transfer programme, be rationed so that a limited number of authorities will be allowed to form companies in any one year.

The process of maintaining and upgrading the fabric of owner-occupied properties is a topic which the previous government largely neglected. Indeed, through restricting the availability of improvement grants it arguably made matters worse. This is an area that the Green Paper addresses directly and canvasses opinion on a proposed new approach to the improvement grant system that could see an extension to local authorities' role.²⁶ The government's concern for standards in the private rented sector has already led to a commitment to introduce a licensing system covering 'houses in multiple occupation', when parliamentary time allows. Licensing in the wider private rented sector is rejected in the Green Paper as being unnecessarily burdensome.²⁷ Relatively modest proposals to replace the housing fitness standard²⁸ with a health and safety rating scale and to overhaul local authorities' powers to intervene where housing is unfit are presented. As alternative approaches to raising standards, the Green Paper raises the possibility that direct payments of housing benefit might be denied to landlords who provide unacceptable housing and that the housing benefit available to tenants in poor quality properties might be restricted in areas of over-supply/low demand. Little appears to be directed at poor quality properties in high demand areas.

Not only the quality of the existing stock, but also the quality of rental housing management is regarded as in need of improvement. In the private rented sector there are a range of modest suggestions, differentiated between those designed to assist 'good and well intentioned landlords' – such as supporting local landlord accreditation schemes and a 'kitemark' for letting agents, and exploring the scope

23 See S. Wilcox, *Housing Finance Review 1999/2000* (York: JRF, 1999).

24 See paras 7.9–7.34.

25 Para 7.39.

26 See paras 4.46–4.55.

27 Para 5.31

28 Currently Housing Act 1985, s 604.

for RSLs to provide low-cost management services – and those designed to ‘make the worst landlords better’.²⁹ In the latter category are the denial of housing benefit payments direct and limited licensing focused on defined local areas of declining demand. Making housing benefit conditional on adequate property and management standards is rejected. On the issue of dealing with harassment and unlawful eviction the government adopts the position that it will review the need to reform the law in the light of research in progress.³⁰ On the other hand, the Green Paper says little new about raising standards in the social rented sector because much of the burden for achieving this is placed upon the existing Best Value regime, the Housing Inspectorate and the Housing Corporation.

A range of more specific suggestions is directed at raising the quality of processes and products in various parts of the housing market. There is a concern to raise the quality of new building, to modernise construction processes and to introduce a Quality Mark to combat cowboy builders in the repairs and maintenance industry. The Green Paper also argues that the problems with take-up of mortgage payment protection insurance flow in part from concerns about the quality and value of the products on offer. It simply reports the ongoing work with lenders and insurers to improve such products.³¹ A new scheme to deal with disputes between landlords and tenants over the return of deposits in the private rented sector is proposed.³² Many of these proposals fall under the broad heading of consumer protection, which the government sees as one of its key contributions to facilitating the operation of the housing market.³³

While choice occupies a prominent position in the Green Paper, it is focused more narrowly. There are some specific issues relating to the private sector – such as the government’s drive to increase the intelligibility of mortgage products to allow consumers to make better informed choices – but the main concern, as it was for the previous government, is to facilitate choice in the social rented sector.³⁴ The discussion is predicated upon the notion of tenant as consumer and choice – between housing providers as well as within the stock of a single provider – is seen not just as desirable but as fundamental to reducing the stigmatisation and reversing the residualisation of social housing. In terms of choice between the dwellings in a social landlord’s stock, the key routes are through potentially far-reaching reforms of the allocations and transfer systems. This is represented by a shift from ‘allocation’ to ‘letting’, and through restructuring rents to create ‘fairer’ patterns of rents. On choice between landlords there is a desire to see a range of social housing providers in a locality ‘competing for tenants’ custom’.³⁵ Where local authorities are transferring stock they should break that stock up to ensure that new monopoly providers are not created. When new build is planned it is not envisaged that this will be local authority stock and hence it will act as a further boost to the RSL sector. Another important, but perhaps more indirect, means of facilitating choice between landlords is the proposal that there should be convergence in rents between neighbouring local authorities and the RSLs operating within their area. The thrust of these proposals is entirely in accord with

²⁹ Paras 5.8–5.25 and 5.25–5.49 respectively.

³⁰ Subsequently published as A. Marsh, P. Niner, D. Cowan, R. Forrest and P. Kennett, *Harassment and Unlawful Eviction of Private Rented Sector Tenants and Park Home Residents*, (London: DETR, 2000).

³¹ Paras 4.6–4.16

³² See paras 5.16–5.17.

³³ Para 4.4.

³⁴ Ch 9.

³⁵ Para 6.14.

that of the previous administration. Indeed, in some respects it takes the positions adopted by the late Thatcher and Major governments a step further.

Much of the discussion in the Green Paper can be linked to a concern with quality or choice, but it also addresses a range of issues which are only loosely related to quality and choice or which appear directed at broader or somewhat different concerns. Some of this discussion reprises existing government initiatives, but there are fundamental questions raised and potentially important proposals advanced.

A theme which vies with quality and choice for prominence in the Green Paper is the role of the local authority in local housing markets. The government indicates its desire to see local authorities take a stronger strategic role which encompasses concern for all tenures.³⁶ The strategic role starts with identifying need and then encompasses a number of strands including facilitating, often through partnership working, the various routes through which needs can be met. Authorities are also in a position to link housing needs to wider social, economic and environmental concerns. Much of the subsequent discussion about improvement in quality contains reference back to the strategic role of local authorities in effecting change. During the course of the twentieth century the role of local authority housing – latterly social housing more broadly – has been transformed from a tenure of choice to a residualised tenure of last resort. In a housing system predicated upon the assumptions that most aspire to own their home and that the private rented sector has virtues of flexibility, there has been limited reflection upon precisely what role social housing can and should fulfil. The Green Paper sets out its own vision, but it also seeks to engender a debate about the future of social housing.³⁷

In the light of existing documents on neighbourhoods and social exclusion originating from the Social Exclusion Unit, the Green Paper addresses issues of the sustainability of communities and of low demand or unpopular housing. These broader debates encompass issues of local economic, crime and regeneration policy rather than simply being housing policy issues, but they clearly inform policy proposals in the social and private rented sector. In the social sector, the reform of allocations policies will not only be about facilitating choice for individuals but also about 'local lettings' policies that aim to change the social composition of areas in order to bring stability and build sustainable communities. The Green Paper also signals existing policies and powers on anti-social behaviour as being a key element in the process of creating and supporting sustainable communities.³⁸ In the private rented sector the Green Paper recognises that it is not simply anti-social behaviour by tenants but the 'large-scale operations of some unscrupulous landlords'³⁹ that are undermining the social fabric of particular localities.

Many commentators were looking for, and expecting the Green Paper to signal, major changes in the way the Housing Benefit system operates because its current form is seen as embodying a whole range of perverse incentives and inefficiencies.⁴⁰ While the Green Paper raises the novel possibility of using Housing Benefit to improve quality in the private rented sector, it does not propose its fundamental restructuring. Instead, at this stage, the government seems content to focus upon improving the administration of the system, reducing the complexity of the rules,

³⁶ Ch 3.

³⁷ See ch 6, also paras 9.11–9.12.

³⁸ Paras 12.27–12.31.

³⁹ Para 5.32.

⁴⁰ See para 11.4; also, for example, P. Kemp, *Housing Benefit: Time for Reform* (York: Joseph Rowntree Foundation, 1998).

tackling fraud and error, and improving work incentives. It signals that longer term fundamental reform is needed, but that the way forward needs to be debated. While some of the initiatives in the Green Paper are to be welcomed, they are at best incremental and there must be concern regarding how existing measures to eradicate fraud are affecting innocent tenants.⁴¹ Proceeding with more comprehensive or extensive anti-fraud measures needs to be considered carefully as not only could they have potentially damaging effects upon people's lives but by underlining the fact that welfare recipients are viewed as actual or potential 'knaves'⁴² it can contribute to the further stigmatisation of those who seek assistance.

Specific Issues

The repositioning of the state

Local authority involvement in housing started with concerns about public health and the direct provision of housing did not become a significant part of that involvement until the early twentieth century. The housing-related activities of local authorities have always been broader than simply providing council housing.⁴³ Yet, in the mid-1980s the idea of the 'enabling' state gained currency and has since risen to dominate policy discourse.⁴⁴ Extended justifications for the view that it was not the state's responsibility to provide but to enable others to do so emerged in the 1990s. But it is clear that commentators operate with different views of what constitutes the 'enabling' role. In particular, those on the political right have tended to see the enabling local state as a minimal 'contractual' state taking a back seat to the private sector in meeting housing need and as engaged in a residual set of activities. Those on the left, in contrast, have tended to interpret enabling as a more strategic role requiring top-down co-ordination of local housing systems.⁴⁵ Goodlad usefully identifies three dimensions across which views of the enabling role vary: top-down planning versus bottom up consumer control; unrestricted state action versus free market distribution; and the state acting alone or acting with others.⁴⁶ The Green Paper sets out an enabling role for local authorities: how is it oriented across these dimensions?

It appears that the Green Paper shifts the emphasis of enabling towards a more top-down orientation, with local authorities expected to assess needs locally and produce a cross-tenure strategy. They are expected to play a role in 'identifying, co-ordinating and facilitating all the resources and agencies that can contribute to the delivery of the strategy'.⁴⁷ Yet this top-down orientation is tempered by the view that '[t]enants, residents, housing associations, private sector landlords, housebuilders, voluntary sector agencies, black and minority ethnic community representatives,

41 For example, Marsh et al, n 30 above, report one senior officer remarking that 'the Rough Sleepers Initiative had been taking people off the streets while the [housing benefit] verification procedures had been putting them back on again' (para 3.74).

42 J. Le Grand, 'Knights, Knaves or Pawns? Human Behaviour and Social Policy' (1997) 26 *Journal of Social Policy* 149.

43 See R. Goodlad, *The Housing Authority as Enabler* (Coventry: IoH/Longman, 1993).

44 The key policy document is DoE, *Housing: The Government's Proposals*, Cm 214 (HMSO, 1987).

45 See D. Osborne and T. Gaebler, *Reinventing Government* (New York: Plume, 1992); N. Ridley, *The Local Right - Enabling not Providing* (London: Centre for Policy Studies, 1988); *Inquiry into British Housing*, (London: NFHA, 1985); R. Goodlad, 'Conceptualising Enabling: The Housing Role of Local Authorities' (1994) 20 *Local Government Studies* 570.

46 Goodlad, n 43 above, 33-39.

47 Para 3.6.

parish and town councils, planning departments, health authorities, social services and the police ... will need to be fully involved in the production of a housing strategy'.⁴⁸ Hence, extensive partnership working is envisaged as a component of policy formulation. Moreover, there is the suggestion that local authorities need to work in partnership across boundaries at the sub-regional level. This partnership theme is also central to consideration of the third dimension of enabling – acting alone or with others – because it is clear that local authorities will be heavily reliant on others to deliver on their strategy. This is particularly the case once the authority has transferred its stock to an alternative RSL landlord. In such circumstances the influence that local authorities have over the nature of the development in their area can at present be limited. It is therefore interesting that the Green Paper proposes that '[w]here the authority is no longer directly responsible for the management of the social housing stock, we would like them to have an increased role in the [Housing] Corporation's decisions on scheme selection and allocations and in monitoring the performance of registered social landlords in their area'.⁴⁹ This implies a shift in the balance of power towards the local authority.

As regards the third dimension of enabling identified by Goodlad – unrestricted state action or free market distribution – there is little in the Green Paper which challenges the direction that policy has taken in the recent past: the growth of home ownership and of voluntary sector not-for-profit housing at the expense of public housing. There is unlikely to be substantial local authority new build. Yet, the emergence of arms-length management companies has the potential to slow the flow of stock from the public to the RSL sector if it establishes itself as a viable alternative.

There are key elements of the Green Paper's agenda such as the view that the separation of strategy from provision within local authorities has not yet gone far enough which would sit comfortably with the Conservatives. The consensus around the wisdom of this sort of separation is intriguing. It ignores most of what we have learned about the nature of the policy process. To attempt to separate strategy from its implementation is to draw an artificial distinction.⁵⁰ Where the two functions are located in different organisations it can lead to confusions and problems of accountability.⁵¹ If the separation is strictly enforced then it reduces the scope for valuable feedback and learning between those who are providing services to those who are shaping the overall direction of activity. Moreover, the requirement to negotiate and liaise with either partner or contractor organisations means that the costs of co-ordinating activities almost inevitably increase. There is limited concrete evidence that the efficiency gains from the creation of smaller-scale, single-purpose housing organisations outweigh the costs of managing and working across the organisational boundaries created. It is not clear that placing provision with an independent provider makes the task of the local authority in delivering a local strategy to meet local housing needs easier or more difficult. The provider organisations can, and in some cases do, legitimately pursue a different agenda once they are independent.⁵² Hence, the benefits of such fragmentation are not unambiguous.

48 Para 3.15.

49 Para 3.17.

50 J. Stewart, 'A Dogma of Our Times – The Separation of Policy-Making and Implementation' (1996) *Public Money and Management* 33.

51 See, for example, K. Walsh, *Public Services and Market Mechanisms* (Basingstoke: MacMillan, 1995).

52 See D. Mullins and M. Riseborough, *Changing with the Times: Critical Interpretations of the Repositioning of Housing Associations*, Occasional Paper 12 (Birmingham: School of Public Policy, University of Birmingham, 1997).

Similarly it is worth reflecting upon the major restructuring of provision away from the state through the large-scale voluntary transfer (LSVT) process. The government uncritically accepts the success of the stock transfer programme and assumes that the experiences of past transfers are a good guide to the wisdom of pursuing the programme further. Yet, early transfers occurred among smaller, often rural, district councils with a relatively low debt burden.⁵³ These authorities managed a housing stock that represented a valuable asset against which private lenders were willing to lend in order to facilitate the transfer. As the programme has extended it has started to encompass urban areas, areas with problems of poor quality stock, and authorities with considerable outstanding debt associated with their stock. The government has recognised the issue through the introduction of Estate Regeneration Challenge Funding and, latterly, one-off special payments to help redeem the debt.⁵⁴ While such payments may facilitate transfer, they represent no additional investment in the stock and no increase in provision. One might ask at what point the extra funding required to make the numbers stack up becomes so great as to render the whole transfer process questionable. Are there a set of authorities for whom the transfer process would represent such poor value for public money that an alternative strategy would be more appropriate? Or are the perceived benefits of leaving the public sector so great that all should be allowed to do so? There is a danger that transfer is taken as a panacea for all ills – a remarkable position for a policy which primarily began simply as a pragmatic attempt to evade the stringent public sector borrowing rules.⁵⁵ It is also the case that transfer is not a foregone conclusion – tenants must vote in favour of transfer. In the past, there has been successful mobilisation of tenants against transfer, particularly as a result of fears about rent rises after transfer.

The government offers a broad vision of local authorities as the key strategic actor at local level. But does the Green Paper seek to ensure that authorities have the necessary powers and tools available to execute that role effectively? There are some suggestions for additional powers – such as dealing with disrepair in the owner-occupied sector – but, in the main, the burden of delivering the strategy rests on the planning framework, authorities' involvement in determining the location and nature of development of new RSL housing through the Approved Development Programme (ADP), and enforcement action that can be taken on unfit properties and anti-social behaviour.⁵⁶ These enforcement powers may be modified to give authorities more leverage over, for example, bad landlords. However, we should question whether these powers are adequate. Experience in using the planning system to deliver affordable housing or other broad housing objectives suggests that it is not a particularly effective tool for meeting housing need. Greater influence over the ADP may be helpful, but is the ADP of an adequate scale to meet needs locally?⁵⁷ Similarly, are regeneration and related funds available to create the sort of physical environments within which sustainable communities can be established? Finally, as we discuss below, the literature on enforcement across a range of policy areas suggests that its effectiveness in raising standards and gaining compliance is, at

53 D. Mullins, P. Niner and M. Riseborough, *Evaluating Large Scale Voluntary Transfers of Local Authority Housing* (London: DoE, 1995).

54 Para 7.29

55 See D. Mullins, P. Niner and M. Riseborough, 'Large Scale Voluntary Transfers' in P. Malpass and R. Means (eds), *Implementing Housing Policy* (Buckingham: Open UP, 1992).

56 Para 3.25

57 The emphasis on the ADP has been backed with increased resources. The subsequent 2000 Comprehensive Spending Review indicated that the ADP will increase from £691 million in 2001–02 to £1.2 billion by 2003–04 (<http://www.housing.detr.gov.uk/csr2000/index.htm> 27 July 2000)

the very least, not to be taken for granted. Overall, therefore, questions regarding the adequacy of the tools to deliver remain.

Central-local relations

Much of the Green Paper's discussion of the role of local authorities is driven by the vision of a strategic authority. As such it suggests that local authorities will have a freedom to develop responses sensitive to local needs and issues. Moreover, the government indicates that in some policy areas – such as planning for new affordable housing⁵⁸ – decision making will be devolved further. Indeed the Green Paper states that 'we have increased the resources and discretion available to local authorities in the delivery of housing policies and investment'.⁵⁹ While this may be the case, local authorities' role needs to be set within the wider context of the regulatory regime within which they operate and the incentives that they face. Over the period since 1980 there have been moves to the devolution and fragmentation of provision, but these have been accompanied by an increase in centralised monitoring and surveillance: restructuring that has been termed 'centralising decentralisation'.⁶⁰ The move to the 'audit society'⁶¹ and a strong evaluation culture has been most evident in the (former) public sector. With the introduction of Best Value and the establishment of the Housing Inspectorate, New Labour have further extended surveillance by the centre of both local government⁶² and the independent RSL sector. Funding and flexibility to act is increasingly being linked to satisfactory performance within this regime and hence the latitude for local actors to follow local priorities is constrained.

One of the most explicit examples of the attempt by central government to increase control over local activity is the discussion of rent policy in Chapter 10 of the Green Paper. The progressive withdrawal of general subsidy to local authority housing since 1980 was the key mechanism by which central government effected the switch in policy from bricks and mortar subsidies to a reliance on housing benefit to assist low-income households. In the RSL sector the mechanism used was the reduction of capital grants for new construction, which required the greater use of private loans. Hence central influence over average rent levels is well established.⁶³ Yet, central government has generally not concerned itself with the method by which social landlords set their rents nor with the rents of individual dwellings. The exception is the Housing Finance Act 1972 which specified a single rent-setting method for all local authorities that would determine the rent on individual dwellings. The Act met with opposition at the local level for its attack on local autonomy to determine policy priorities⁶⁴ and was repealed in 1974. It was not until the 1989 Local Government and Housing Act that central government

58 See para 8.2.

59 Para 3.30.

60 P. Hoggett, 'New Modes of Control in the Public Service' (1996) 74 *Public Administration* 9.

61 See M. Power, *The Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1997).

62 See paras 3.26–3.29; also M. Geddes and S. Martin, 'The Policy and Politics of Best Value: Currents, Crosscurrents and Undercurrents in the New Regime' (2000) 28 *Policy and Politics* 379 for an overview.

63 See P. Malpass, *Reshaping Housing Policy: Subsidies, Rents and Residualisation* (London: Routledge, 1990); J. Hills, *Unravelling Housing Finance* (Oxford: Clarendon Press, 1991); B. Walker, 'Incentives, Choice and Control in the Finance of Council Housing' in A. Marsh and D. Mullins (eds), *Housing and Public Policy* (Buckingham: Open UP, 1998).

64 Most notably by the councillors of Clay Cross Urban District Council: P. Malpass, 'The Road from Clay Cross' in C. Grant (ed) *Bullt to Last? Reflections on British Housing Policy* (London: Roof, 1992).

again attempted to influence the way that rents of individual properties were set, ostensibly to combat perceived problems of overconsumption and encourage the more efficient use of stock. This attempt was largely unsuccessful.⁶⁵

In this context the Green Paper represents a potentially significant development because it proposes that one of three options should be used to determine the rents for classes of property. It could be interpreted therefore as a move towards a system in which central government seeks to remove local discretion and effectively determine the rents on individual dwellings. All three options are based to some extent on capital values. For many local authority landlords, implementing any of these options would represent a major upheaval because very few base their rents on capital values at present and most of those who have changed their rent-setting method recently have changed to points-based systems.⁶⁶ The proposals in the Green Paper are not restricted to the local authority sector. They encompass both the local authority and RSL sectors and indicate that the government would like to see rents for similar properties in the two sectors converge to similar levels over a period of ten years. Convergence would be achieved by restricting the rate of increase of (currently higher) RSL rents below the level of local authority rents. However, if an RSL's business plan is thrown into question by the proposed rent cap then it appears that it will be allowed to operate outside the regime: hence convergence will be partial at best.

The rationale for the policy is the desire to create 'fair' rent-setting systems in which people pay more for greater amenity and residents of similar properties owned by different landlords do not face markedly different rents. Modelling of the options at regional level and of capital value systems at local level suggests that there could be large changes in rents in some areas and for some types of property.⁶⁷ Fairer systems of rent setting are seen as removing artificial distortions from tenants' choice about where to live and from whom to rent their home. As such, the policy seems inspired by a rather straightforwardly economic view of choice behaviour which treats housing choices as similar to buying any other good. Whether such a model can sensibly be applied to housing choices needs careful consideration. There is some evidence that changes in rent levels have only limited impact upon local authority tenants' choice of dwelling and that not all tenants share the government's view on what constitutes a 'fair' rent system.⁶⁸

Equally importantly, the Green Paper sets out the government's objectives in rent setting but it has less to say about how these objectives might be realised. Particularly with respect to RSLs, at present central government has very few effective levers with which to bring about the type of changes envisaged. Whether any subsequent legislation will include the creation of new mechanisms for influence is by no means clear. An interesting question is how RSLs in particular – as independent bodies – will react to this increased direction from the centre. While rent policy is an area where centralisation is proposed, it is also an area in which the centralising and decentralising tendencies of New Labour come into conflict.

65 See B. Walker and A. Marsh, *Rent Setting Policies in English Local Authorities* (London: DoE, 1995); B. Walker and A. Marsh, 'Pricing Public Housing Services. Mirroring the Market?' (1998) 13 *Housing Studies* 549.

66 See B. Walker and A. Marsh, *Rent Setting Policies in English Local Authorities, 1998/99* (London: DETR, 2000).

67 HACAS Consulting, *Impact and Implications of Restructuring Rents in the Registered Social Landlord Sector: Summary* (London: DETR, 2000); A. Freeman, A. Holmans and C. Whitehead, *Evaluating Housing Affordability* (London: CIH/LGA/NHF, 1999).

68 See, for example, B. Walker, P. Niner, M. Wardman and A. Marsh, *Social Housing Tenants' Attitudes and Reactions to Rent Levels and Rent Differentials* (London: DETR, 2000).

The government strongly supports tenant consultation and participation in the social rented sector and is championing the notion of Tenant Participation Compacts.⁶⁹ Compacts cover a range of activities and are intended to include rent-setting policy. Yet if central government plans to shape rent setting are implemented then there will be no local process in which tenants can participate. How this tension will be resolved remains to be seen.

Discretion

Much of the early literature on the allocation of social housing drew attention to the problems created by the exercise of 'discretion'.⁷⁰ In the early literature, discretion was 'depicted as the bug in the system – a source of deviance which allowed short-term management goals to compromise the principle of social justice.'⁷¹ It was regarded as a principal cause of discrimination in housing allocation.⁷² Inconsistency between local authorities was regarded as so problematic in 1989 that, despite acknowledging differential supply of accommodation between areas, an amended Code of Guidance was suggested to hem in discretion under the homelessness legislation.⁷³ The dichotomy between discretion and rules, familiar to administrative lawyers, though probably false,⁷⁴ has nevertheless proved important for housing professionals who have sought to define their processes more clearly. In particular, local authorities have been sensitive to criticism that their policies, procedures and decision-making are productive of racial disadvantage and discrimination.⁷⁵ Some have thus sought to redraw the boundaries in order to combat this problem.⁷⁶

Until 1996, legislation barely acknowledged that local authorities had a housing stock: as for selecting households and allocating the stock to them, the legislation prescribed four 'reasonable preference' categories which had remained the same since 1935.⁷⁷ In allocating their stock, most local authorities adopted schemes which awarded points to applicants on the basis of pre-defined criteria.⁷⁸ These schemes were attempts to design-out discretion from the system as much as possible.⁷⁹ In the mid-1990s, some of these schemes became so complex and multi-layered that some level of change became almost inevitable. Indeed, it became acknowledged that points schemes contained as many problems of discretion as previously.

69 See <http://www.housing.detr.gov.uk/information/compacts/index.htm>.

70 N. Lewis, 'Council Housing Allocation: Problems of Discretion and Control' (1976) *Public Administration* 147;

71 S. Smith and S. Mallinson, 'The Problem with Social Housing: Discretion, Accountability and the Welfare Ideal' (1997) 24 *Policy and Politics* 339, 341.

72 See generally, J. Morris and M. Winn, *Housing and Social Inequality* (London: Hilary Shipman, 1990).

73 See DoE, *The Government's Review of the Homelessness Legislation* (London: DoE, 1989) para 64.

74 See R. Sainsbury, 'Administrative Justice: Discretion and Procedure in Social Security Decision-Making' in K. Hawkins (ed), *The Uses of Discretion* (Oxford: OUP, 1992).

75 See the various investigations of the Commission for Racial Equality in this area, especially *Race and Council Housing in Hackney* (London: CRE, 1984).

76 Apparently with little success – see the important work of S. Jeffers and P. Hoggett, 'Like Counting Deckchairs on the Titanic: A Study of Institutional Racism and Housing Allocations in Haringey and Lambeth' (1996) 10 *Housing Studies* 325; and S. Halliday, 'Institutional Racism in Bureaucratic Decision-Making: A Case Study in the Administration of Homelessness Law' (2000) 27 *Journal of Law and Society* 449.

77 Housing Act 1985, s 22; judicial decisions actually limited any such preference further: *R v Wolverhampton MBC ex parte Watters* (1997) 29 HLR 931.

78 See P. Prescott-Clarke et al, *Routes into Local Authority Housing* (London: DoE, 1994).

79 As suggested in the Cullingworth Committee report, *Council Housing: Purposes, Procedures and Priorities* (London: HMSO, 1969).

This agenda has gradually been replaced by a more pressing (financial) concern that some local authority and RSL estates are so unpopular that no household wishes to live there. The current explanation for this draws on a simplistic framework that '[t]he most important factors across all sectors are the poor perception or stigma of areas and crime and anti-social behaviour. Poor quality environments and the design and type of housing or estates are of moderate importance along with overall housing surplus.'⁸⁰ Other factors, such as demography and increasing marginalisation of the occupants of social housing, are also conventionally discussed in this context.⁸¹ The loss of revenue from empty social housing, in a climate of risk of bankruptcy for RSLs and fiscal crisis for local authorities, cannot be sustained.

Whilst the Green Paper retains the commitment that 'priority for social housing should generally continue to be given to people in the greatest housing need', an important caveat is made for situations 'when it is necessary and desirable, for some wider community benefit, to allow exceptions to this.'⁸² The caveat refers to 'local lettings' schemes – through which allocators will be entitled to 'correct a significant social imbalance such as an excessive child density on certain estates.'⁸³ In these areas, meeting housing need – the foundational principle of social housing – is irrelevant; rather the concern is to engineer 'sustainable communities', the current policy priority.

What constitutes a sustainable community is left ultimately to the discretion of the local authority. Indeed, the whole framework of social housing allocations – as distinct from the selection of households – will return to broad discretion.⁸⁴ Shifts to advertising units of social housing, as opposed to allocating them, replaces one discrete decision with layers of further discretion, not to mention changing the focus of applicants.⁸⁵ On one level, given the apparent regional and national differences in supply and demand for social housing, this was the only rational legislative (non-) framework. Yet it re-opens the debate about the value of discretion (quite apart from rendering local allocations schemes more difficult to challenge through judicial review), which becomes much more apparent and less structured than ever before. Given the problems already outlined, which were said to be at least partially due to the existence of discretion, there are clear problems in these proposals.⁸⁶ Equally, the new system enables social housing providers to create for themselves categories of 'more eligible' households. This is dangerous partly because certain groups already marginalised will be further marginalised in this way, and partly because it privileges those who have the requisite knowledge and ability to bargain. Finally, it is not clear that it is possible to produce the desired end-product.⁸⁷ Attempts to

80 Policy Action Team 7, *Report by the Unpopular Housing Action Team* (London: SEU, 1999), para 2.44; cf S. Damer, 'Wine Alley: The Sociology of a Dreadful Enclosure' (1974) 22 *Sociological Review* 221; P. Watt and K. Jacobs, 'Discourses of Social Exclusion' (2000) 17 *Housing, Theory and Society* 14.

81 See the influential report of D. Page, *Building for Communities* (York: Joseph Rowntree Foundation, 1992); cf I. Cole, G. Gidley, C. Ritchie, D. Simpson and B. Wishart, *Creating Communities or Welfare Housing* (Coventry: Chartered Institute of Housing, 1996).

82 Para 9.12.

83 Para 9.29.

84 cf Housing Act 1996, Part VI.

85 The model on which it is based, derived from Delft in Holland, may not produce the desired results: see J. Kullberg, 'From Waiting Lists to Adverts: The Allocation of Social Rented Dwellings in the Netherlands' (1997) 12 *Housing Studies* 393.

86 Whether or not these problems can be put down to discretion on its own, rather than broader structural deficiencies in the social housing system such as budget cuts and personnel efficiency gains (which themselves will not be remedied), is itself open to doubt – see Halliday, n 6 above.

87 Indeed, the claims for the end-product are so high that it is unlikely they would ever be met anyway.

revive the sector are bound to fail if they do not address the root causes of the sector's unpopularity. This would in part require a shift in rhetoric away from eulogising the private sector, denigrating the social sector and local authority housing management itself.⁸⁸

Rights and responsibilities

Many of the twentieth-century struggles around housing policy concerned the development of rights for the occupants of housing: Glasgow rent strikes led to the development of security of tenure and rent control in the private rented sector; mortgage strikes led to regulation of mortgage lenders' relationships with building societies; the squatters movement and other campaigning work, such as the Ken Loach film, *Cathy Come Home*, led to the homelessness legislation; security of tenure for public sector occupants was finally gained in 1980 after a long-running campaign by, for example, the Consumers' Association.⁸⁹ The development of these rights also necessarily set out, or reinforced,⁹⁰ the responsibilities of households to conduct themselves in an appropriate manner. Those who did not were liable to be evicted and excluded from that tenure. In the latter part of the twentieth century, housing policy came to concentrate more on the responsibilities of occupants.⁹¹ Local authorities were empowered to *exclude* households from their waiting lists indefinitely – explicitly in Housing Act 1996,⁹² although this merely codified the then current practice – and, in particular, anti-social behaviour became a concept with particular political currency in the housing system and beyond.⁹³ Housing policy, in fact, began to adopt and adapt crime control models in an attempt to root out deviance – 'zero tolerance' housing management, for example, became a topic of regular discussion.⁹⁴ The principal problem created by such policies was that, whilst households could be excluded, the 'problem' was passed on to another (probably poorer) sector.

The understanding behind these policies and practices was that breach of the standards of responsibility by households potentially led to exclusion for an unlimited period. The Green Paper, on the other hand, seeks to operate a broader system of rights, penalties *and redemption*. Anti-social behaviour appears often to be the driver behind each of these (although it rarely appears in the Green Paper discussions, as it has been the subject of policy interventions elsewhere).⁹⁵ As regards rights, many will no doubt welcome the suggestion that there is to be a new single tenancy agreement for all social housing occupants, repairing the problematic division which currently exists between local authority and RSL security of tenure.⁹⁶ In setting up this new system, however, it is clear that there will be flexibility for social housing providers to use the assured shorthold tenancy

⁸⁸ See I. Cole and R. Furbey, *The Eclipse of Council Housing* (London: Routledge, 1994) ch 7.

⁸⁹ Respectively the Increase in Rent and Mortgage (War Restrictions) Act 1915; Building Societies Act 1939; Housing (Homeless Persons) Act 1977; Housing Act 1980. As regards the first two, see P. Piratin, *Our Flag Stays Red* (London: Lawrence & Wishart, 1978).

⁹⁰ See E. Burney, *Crime and Banishment* (Winchester: Waterside Press, 1999).

⁹¹ See Cowan, n 3 above, Part III.

⁹² s 161(4).

⁹³ See, for example, S. Scott and H. Parkey, 'Myths and Reality: Anti-Social Behaviour in Scotland' (1998) 13 *Housing Studies* 325; C. Hunter, J. Nixon, and S. Shayer, *Neighbour Nuisance, Social Landlords and the Law* (Coventry: Chartered Institute of Housing, 2000).

⁹⁴ See, for example, S. Harriott, 'Zero Tolerance in Housing' (1998) June/July *Housing Review* 28; N. Malik, 'Nowhere to Run' (1998) September/October *Roof* 16.

⁹⁵ See, for example, the Social Exclusion Unit's Policy Action Team 8, *Anti-Social Behaviour* (London: SEU, 2000).

⁹⁶ Paras 9.66–9.72.

mechanism to let property at market rents to those who are not in housing need. Thus, the creation of 'sustainable communities' – the holy grail of housing policy – concerns mixing poor with not-so-poor households and the abandonment, or at the very least the suspension (when required), of the universalist notion of need.

Rather than automatic exclusion from the local authority sector (and probably most other tenures as well), the Green Paper's proposal is for a rather different system of punishment which offers the chance for redemption (on the prison-parole-release model). The previous system of exclusions did local authorities no favours – first, the sector began to operate in an increasingly reactionary way and, secondly, those excluded were often those traditionally regarded as being in housing need⁹⁷ – and RSLs often operated exclusions without explicit authority or publicity.⁹⁸ The suggested punishments are a temporary reduction in waiting list priority or a temporary suspension from the waiting list – 'people would be given a clear indication of what they needed to do to get the restriction lifted' – only if the circumstances of the household allow this (that is, they are not too vulnerable).⁹⁹ In this case, redemption is conditional.

Another example of this approach is in the possible new arrangements regarding housing benefit payments. The problem is set up in the following way: 'A handful of anti-social [private rented sector] tenants can have a devastating impact on the social fabric of the surrounding area. The problem is compounded by the fact that anti-social tenants are often dependent on housing benefit, paid directly to the landlord.'¹⁰⁰ It is suggested that housing benefit could be 'adapted to encourage both tenants and landlords to behave responsibly'.¹⁰¹ Two methods are canvassed: first, reductions in housing benefit for 'unruly tenants' (to be defined by an objective test), and second, reductions for unruly tenants as part of a process of pursuing an anti-social behaviour order. Redemption occurs (that is, housing benefit is restored to its full allowable amount) 'once the tenant's behaviour ha[s] improved'.¹⁰² There must also be some concerns as to the compatibility of such a proposal with the Human Rights Act 1998 if housing benefit is to be seen as a civil right.¹⁰³ Although this is not a dramatic shift in housing policy, it does represent a mollification of the previous strategy. Most of that previous strategy nevertheless will remain in place, with the concomitant intractable problems facing housing managers, and, indeed, has been enhanced by the penal nature of New Labour's crime and disorder policy.

Control/persuade

For some time the housing system, as other policy systems, has been operated under informal codes of guidance, circulars, and other extra-statutory mechanisms. Methods of controlling actors in the housing system through primary legislation have often had effects and impacts which were not anticipated and/or desired.¹⁰⁴

97 For general discussion, see D. Cowan, R. Gilroy and C. Pantazis, 'Risking Housing Need' (1999) 26 *Journal of Law and Society* 403.

98 See, for example, D. Cowan, C. Pantazis and R. Gilroy, M. Bevan, *Housing Sex Offenders: An Examination of Current Practice* (Coventry: Chartered Institute of Housing, 1999).

99 Paras 9.13–9.15.

100 Para 5.4.

101 Para 5.46.

102 Para 5.47.

103 We are grateful to an anonymous referee for this point.

104 See, for example, M. Partington, 'Rethinking British Housing Law: The Failure of the Housing Act 1988' in M. Freeman (ed), *Critical Issues in Welfare Law* (London: Stevens & Sons, 1991).

Policies such as compulsory competitive tendering, which adopted a command and control structure, have since been substituted by the more open-textured system of Best Value.¹⁰⁵ The Green Paper sets clear limits to central government's involvement in housing regulation in certain crucial areas, preferring to leave such regulation to best practice initiatives, voluntary codes of regulation on the part of umbrella organisations, and voluntary licensing systems designed to increase standards. Thus, those with a bent towards autopoietic regulatory strategies will find much of interest.

Such strategies are most in evidence when the Green Paper deals with the private sector, especially renting. Access to owner-occupation is the point at which the persuasion form of regulation bites – so, for example, the government will set 'benchmark standards on charges, access and terms', has delegated regulation of mortgage selling to the Financial Services Authority, and entered into partnerships with insurers concerning the mortgage payment protection insurance package.¹⁰⁶ All of this is meant to complement the Council for Mortgage Lenders' own voluntary code.

As regards private renting, the Green Paper outlines the problem as being that the sector has 'a far worse image than it deserves'.¹⁰⁷ There are a number of reasons why this image has developed – chief amongst these have been the repercussions of accusations of Rachmanism from which, arguably, it has not hitherto been able to pull back.¹⁰⁸ Properties are commonly regarded as being over-priced and poor quality, landlords' management poor and exploitative, and households in the sector as desperate to leave or stuck in it. Regulation is not an option because it is regarded as stifling the sector (a point to which we return below):

In short to raise [small landlords'] standards and prosper, they need encouragement, support and education rather than further heavy regulation. This can best be provided by local authorities, professional and voluntary organisations, though we are anxious to play our part by helping to spread best practice and provide general support.¹⁰⁹

The proper role of central government, therefore, is constructed as being strategic and facilitative. Voluntary licensing and landlord-accreditation schemes, management of tenants' deposits, and a letting agents' kitemark developed by the agents' governing body are proper matters for the industry itself, supported by local government.¹¹⁰ Central government facilitates through its research priorities. Suggestions that the whole sector – rather than just 'houses in multiple occupation' – should be subjected to a compulsory licensing scheme are rejected on the basis that this would be 'a massive undertaking' and involve too much 'red tape [for] some perfectly respectable landlords' who would exit the sector.¹¹¹

This only serves to highlight just how privileged the private rented sector is within the housing system, in contrast to the social rented sector. Gone are the days when commentators were arguing that 'the provision of dwellings for renting is no

¹⁰⁵ Vincent-Jones, n 4 above.

¹⁰⁶ See paras 4.6–4.11.

¹⁰⁷ Para 5.5; for discussion, see A. Crook and P. Kemp, 'The Revival of Private Rented Housing in Britain' (1996) 11 *Housing Studies* 51.

¹⁰⁸ D. Nelken, *The Limits of the Legal Process – A Study of Landlords, Law and Crime* (Edinburgh: Academic Press, 1983) is prefaced with the following definitions of Rachmanism: 'systematic intimidation of tenants to extract exorbitant rents' (Penguin Dictionary); 'conduct of a landlord who charges extortionate rents for property in which very bad slum conditions prevail' (Chambers Dictionary).

¹⁰⁹ Para 5.10.

¹¹⁰ Paras 5.11–5.17.

¹¹¹ Para 5.31.

longer an appropriate activity for private enterprise';¹¹² the current position, shared by all political parties, is that the sector must be revived and is urgently required as the stepping stone leading to (or from) owner-occupation. Thus, in this context, quality, a theme of the Green Paper itself and a problem in the sector, is regarded as a problematic largely outside the control of government.

Unanswered questions

Disarming government

In no sense can the Green Paper be regarded as even the start of a root-and-branch revision of housing policy. At its heart lies a series of assumptions and received wisdom concerning the housing system which are not touched upon, let alone questioned. At the centre of housing policy, they are debilitating because they severely delimit the opportunities for reform. Most significantly, this can be seen in the discussion of the private rented sector. It is assumed that in order for that sector to grow, deregulation and its spirit are necessary. Almost since they were introduced, security of tenure and rent control (or its successor rent regulation) have been regarded as responsible for the sector's historic decline.¹¹³ The introduction of the assured shorthold tenancies (which last a minimum period of six months and require a minimum of two months notice as well as a court order for termination) and a market rent structure in the Housing Act 1988 have been regarded as responsible for the reversal in the sector's fortunes.¹¹⁴ The Green Paper appears to accept this position without comment. However, such a position may be contested on two fronts: first, it can be argued that the sector's decline was contingent upon growth in the other housing sectors (particularly owner-occupation and local authority), better rates of return in other markets (particularly equity markets), and occurred especially during the post-war slum clearance campaigns;¹¹⁵ second, the revival in the sector was not necessarily due to deregulation but once again contingent upon other aspects of the housing system, especially the development of negative equity and mass mortgage arrears in the early 1990s.¹¹⁶ If one were to draw upon the wealth of available research in this area, it would certainly be possible to challenge the assumptions that deregulation itself has been successful and no further government intervention in the sector is desirable.

A further aspect of the disarming of central government lies in what it is simply unable to do because of the legacy of previous housing policies (which, no doubt, the current government would in any event accept). Policy-making is thus hemmed in by the cumulative effects and impacts of previous policies. A good example of this concerns RSL rent structures. From 1987, it was central government's policy that most allocations of capital to RSLs for development had to be matched by a certain percentage of private finance. Since that time, RSLs have developed more

112 F. Berry, *Housing. The Great British Failure* (London: Croom Helm, 1974) 123

113 For discussion, see D. Cowan, n 3 above, ch 2 and the report of the Francis Committee, *Report of the Committee on the Rent Acts*, Cmnd 4601 (London: HMSO, 1971).

114 See DoE, *The Legislative Framework for Private Renting*, Consultation Paper linked to the Housing White Paper *Our Future Homes* (London: DoE, 1995).

115 Slum clearance particularly affected the private rented sector and shifted tenure to local authorities: House of Commons Environment Committee, *The Private Rented Sector*, HC 40-I (London: HMSO, 1982); more generally see P. Kemp, 'Ideology, Public Policy and Private Rental Housing since the War' in P. Williams (ed), *Directions in Housing Policy* (London: Paul Chapman, 1997).

116 A. Crook and P. Kemp, 'The Revival of Private Rented Housing in Britain' (1996) 11 *Housing Studies* 51.

entrepreneurial attributes. However, this all occurred when their revenue was unaffected by regulation. Although the Housing Corporation was forced into controlling rent rises in the sector because of concerns about affordability and levels of housing benefit¹¹⁷ – this issue has remained particularly intractable because the greater the controls on rents, the less scope there is for repayment of interest on past and future outstanding loans.¹¹⁸ Thus, although quite significant changes are mooted in relation to social sector rents – particularly converging local authority and RSL rents – any such reforms ‘will have to take particular account of the impact on the finances of’ RSLs.¹¹⁹ Although the Green Paper, in rare strong-arm mode, talks of mergers between RSLs, at the heart of the policy of convergence lies an acceptance that convergence may not be possible: ‘[RSLs] are independent bodies. They must be able to set rents at a level that allows them to meet their financial commitments, maintain their stock and continue to function as financially viable organisations.’¹²⁰

Managing governance

Over the 23 years since the 1977 review, the housing system has fractured and fragmented. Organisational responsibility can no longer be located in single bureaucracies, the mortgage market is no longer dominated by building societies (few still exist), and RSLs are scarcely recognisable from their position in 1977. The complexity of the housing system has led to real and understandable concerns about how it can, or should, be managed.¹²¹ Issues about accountability,¹²² policy implementation,¹²³ collaboration and co-operation are ongoing concerns.¹²⁴ The centre of power has shifted to diverse groups and organisations. Multi-agency or partnership working – through policy networks – is a ubiquitous feature of the terrain of housing policy implementation as a response to this fragmentation.¹²⁵

Such a description of the housing system does not fit neatly with the Green Paper. One might have expected to find discussion of the following issues: accountability of RSLs and the Housing Corporation, together with the proliferation of other voluntary housing organisations; the development of common housing registers which seek to give equality of opportunity for households to access social housing in an area and act as a simpler more integrated means of access to fragmented provision; the management of the vast array of housing finance products available to individual households; co-ordination and co-operation between agencies (particularly, for example, taking into account data protection concerns). There is occasional

117 See DoE, *More Choice in the Social Rented Sector*, Consultation Paper linked to the Housing White Paper, *Our Future Homes* (London: DoE, 1995).

118 See, for example, Housing Corporation, *Investigations into Registered Social Landlords Rents 1999/2000* (London: Housing Corporation, 2000).

119 Para 10.32.

120 Para 10.39.

121 See, for example, P. Malpass (ed), *Ownership, Control and Accountability: The New Governance of Housing* (Coventry: Chartered Institute of Housing, 1997);

122 See, for example, the concerns of the Nolan Committee, *Standards in Public Life: Local Spending Bodies*, Cm 3270 (London: HMSO, 1996); P. Day, D. Henderson and R. Klein, *Home Rules: Regulation and Accountability in Social Housing* (York: Joseph Rowntree Foundation, 1993).

123 B. Read, ‘Interorganisational Networks and the Delivery of Social Housing’ (1995) 10 *Housing Studies* 133.

124 K. Jacobs, *The Dynamics of Local Housing Policy* (Aldershot: Ashgate, 1999); C. Oldman, ‘Working Together to Help Homeless People: an Examination of Inter-Agency Themes’ in R. Burrows, N. Pleace and D. Quilgars (eds), *Homelessness and Social Policy*, (London: Routledge, 1997).

125 See generally, R. Rhodes, *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability* (Buckingham: Open UP, 1997).

reference to joint working (as with previous policy documents), but the issues which underlie it are not dealt with. Thus, the Green Paper appears to be out of touch with the way in which the increasingly complex housing system is, and needs to be, managed. In this respect, it is, oddly, Old Labour.

The cost of housing

Throughout the post-war period the overall aims of housing policy have typically coupled the provision of good quality accommodation with some notion of bringing the cost of that accommodation within the means of all. The precise formulation of the policy aim may have varied over time, but the underlying objective has remained in more or less ambitious form. A typical formulation appears in the 1971 White Paper which viewed the goal of housing policy as 'a decent home for every family at a price within their means'.¹²⁶ Whether the 'within their means' element of this policy aim was ever articulated with enough precision or pursued with enough vigour is a point of debate,¹²⁷ but what cannot be denied is that there was some concern with what is now referred to as 'affordability'. The affordability question emerged as a key policy issue in the late 1980s and early 1990s as house prices increased rapidly, private rents were deregulated and the government pursued policies which raised rents in the social rented sector. The government was criticised for failing to provide an explicit definition of what it considered to be 'affordable' housing and for neglecting the poverty trap and work disincentive effects the higher rents policy was creating.¹²⁸

The Green Paper clearly echoes previous policy statements with its subtitle: 'A decent home for all'. But the affordability element of the traditional formulation is absent. Similarly, when we consider the government's eight key principles for housing policy the closest they come to a commitment on housing costs is in the statements 'giving responsibility to individuals to provide their own homes where they can, providing help for those who cannot' and 'reducing barriers to work, particularly in relation to benefit and rent policy'.¹²⁹ Other key principles – such as 'promoting sustainable development that supports thriving, balanced communities' might be taken to imply a concern for housing costs to ensure that lower-income households are not priced out of local housing markets and hence the communities to become 'unbalanced'. But this is not made explicit. Elaboration of these principles brings with it reference to ensuring the provision of 'affordable social housing'¹³⁰ and the subsequent discussion of rent restructuring locates affordability as a key concern. Yet even here it is one among a number of concerns including the impact of rent levels on public expenditure and the 'impact on the finances of social landlords, especially registered social landlords, who must be able to maintain their stock and meet their commitments to lenders'.¹³¹ The only time that 'affordable' occupies a prominent position in the Green Paper discussion is with

¹²⁶ DoE, *Fair Deal for Housing*, Cmd 4728 (London: HMSO, 1971).

¹²⁷ For an extended discussion see J. Hills, 'Housing: A Decent Home within the Reach of Every Family?' in H. Glennerster and J. Hills (eds), *The State of Welfare: The Economics of Social Spending*, (Oxford: OUP, 2nd ed, 1998).

¹²⁸ For discussion see, for example, G. Bramley, 'An Affordability Crisis in British Housing: Dimensions, Causes and Policy Impact' (1994) 9 *Housing Studies* 103; J. Ford and S. Wilcox, *Affordable Housing, Low Incomes and the Flexible Labour Market*, Research Report 22 (London: NHFA, 1994); Freeman et al, n 67 above.

¹²⁹ Para 1.5.

¹³⁰ Para 1.9.

¹³¹ Para 10.3.

respect to the use of the planning system and the supply of affordable new homes.

It appears that assuring accommodation is affordable is given no greater emphasis than it was by the previous administration. It could be argued that it is accorded less prominence than might have been expected. The key to why this might be the case, with respect to social housing at least, is provided by the references to public spending and RSLs' commitments to lenders. As we discussed above, the government is hemmed in by the commitments already made to private lenders. Unless very radical action is contemplated, its room for manoeuvre is heavily constrained. This applies equally to discussions about the reform of housing benefit, where one of the reasons that major change was not proposed was undoubtedly that this would upset the business plans of RSLs. In this instance, however, continuity has arguably worked in favour of tenants in as much as it means that the generosity of the benefit has not been restricted further.

Conclusions: same old Tory housing policy?

Since taking office, the New Labour government has produced a vast array of documents dealing with or touching upon the housing system. The Green Paper brings much of this together within the twin themes of quality and choice. In its themes and some of the proposals, the Green Paper could quite easily have emanated from the Conservative government. Proposals to encourage councils to sell off their stock; concerns about 'perverse incentives' within the social housing allocation and housing benefit systems; rent-setting policies; the focus upon owner-occupiers as consumers; reinvigoration of the private rented sector; voluntary regulation schemes; strategic role of local authorities (rather than as providers of social housing); and concerns about the effects of anti-social behaviour leading to local lettings schemes – all of these were in the foreground of Conservative housing policy. The detail, on the other hand, is rather different. There appears to have been a shift away from command and control frameworks to carrot-and-stick approaches; for example, once-and-for-all exclusions are to be replaced by a policy of redemption and there is to be compulsory licensing of houses in multiple occupation. And there are areas – such as the proposed changes to the improvement grant system – where significant and positive change of direction may occur.

However, the Green Paper highlights a number of contradictions in central government housing policy which will, no doubt, impact upon its implementation. For example, one might question how well local authorities will be able to put in place strategies for dealing with the private rented sector when there is to be more centralisation and central control of that sector. It is also unfortunate that the government did not use the Green Paper as an opportunity to conduct a root-and-branch assessment of housing policy. This is likely to reflect the real constraints on central government's management of the housing system, but arguably it also reflects a reluctance to challenge some of the received wisdom that has now come to dominate and constrain housing policy thinking. Whilst Malpass suggested that the previous government's White Paper had no policies designed to address the housing problems in 1995, New Labour's Green Paper provides only modest or superficial solutions.¹³² The structural problems facing the different housing sectors will remain largely untouched.

¹³² n 11 above.

The Criminal Cases Review Commission: Reporting Success?

*Richard Nobles and David Schiff**

The Criminal Cases Review Commission (CCRC) issued its third Annual Report on 15 June 2000.¹ In it, and in other communications about its work,² the Commission makes claims that it is successful and perceived as successful, and that the doubts that had been expressed about its likely success prior to its establishment have been dispelled.³ After having operated for three years this seems to be an appropriate time to assess the claimed success of this non-departmental public body in carrying out its responsibilities as a post-appeal institution designed to remedy miscarriages of justice.⁴ However any assessment concerning success requires a measure by which that assessment can be made, and such a measure is principally dependent on two factors. First, it involves the need to have a clear idea of what constitutes a miscarriage of justice. And secondly, to assess the role of the CCRC it is necessary to put it into its institutional and non-institutional context. This involves understanding the relationship between this body and other institutions and pressure groups⁵ that are aimed at remedying miscarriages of justice. What are those relationships and what ought they to be? In brief, the question 'what should this institution be seeking to achieve?' is a

* Law Department, London School of Economics. We would like to thank Nicola Lacey for commenting on a draft of this article.

- 1 The first Annual Report, 1997–98 (AR1) covered the period from the Commission's establishment on 1 January 1997 to 31 March 1998. The second Annual Report, 1998–99 (AR2) covered the period from 1 April 1998 to 31 March 1999 and the third Annual Report, 1999–2000 (AR3) the period from 1 April 1999 to 31 March 2000.
- 2 A good example is the presentation of oral evidence by the Commission's Chairman, Sir Frederick Crawford and its Chief Executive, Mrs Glenys Stacey to the House of Commons Select Committee on Home Affairs on 11 April 2000 (*The Work of the Criminal Cases Review Commission*, HC 429 (2000)).
- 3 Sir Frederick Crawford has made this claim in his covering memos to the Home Secretary in each of the Annual Reports with increasing confidence. 'Before the Commission was established, gloomy forebodings were widely expressed about its prospective independence and ability to investigate suspected miscarriages of justice thoroughly. Those doubts have been allayed during the Commission's start-up year.' (AR1, 5) 'I believe that the Commission can be justly proud of what it has achieved, in a difficult and sensitive role, in the short space of two years since its inception.' (AR2, 5) 'Familiarity with its work has largely allayed early stakeholder concerns regarding its independence and its ability to investigate suspected miscarriages of justice thoroughly.' (AR3, 5).
- 4 It has a number of other related responsibilities, as set out in the Criminal Appeal Act 1995.
- 5 These include organisations such as Justice, which has for nearly forty years assisted those claiming to have been victims of miscarriages of justice, and Liberty. Indeed, prior to the establishment of the Commission, Justice considered itself to be 'the main independent body which receives and reviews allegations of miscarriages of justice ... mainly with those in which appeal rights have been exhausted, or where detailed investigations are necessary.' (Justice, *Remedying Miscarriages of Justice* (London, 1994) 21). Pressure groups also include elements of the media, such as the television programmes 'Rough Justice' and 'Trial and Error', as well as individual investigative journalists who have played a significant role in bringing individual miscarriages of justice to the attention of the public. They also include ad hoc committees set up by family and friends of those alleging miscarriage.

necessary and, in our view, more informative preliminary question to that of 'how well is it doing?'⁶ Whilst there is a definite consensus that miscarriages of justice should be remedied, this consensus does not translate into agreement as to what constitutes miscarriage. Neither does it determine what the respective responsibilities are of the different institutions and pressure groups involved or what level of public expenditure should be devoted to dealing with the problem. In this article, we seek to examine the success of this body against the target of a range of attitudes and ambitions that led to its creation, as well as the ambitions that it claims for itself.⁷ The main sources of information on which we rely are the Commission's Annual Reports.⁸

The task facing the Commission

Setting up the Commission was a leading recommendation of the Royal Commission on Criminal Justice,⁹ a body whose own creation was announced on the day that the Birmingham Six appeal was upheld and their convictions quashed in 1991. That successful appeal followed the earlier successful appeal by the Guildford Four in 1989, and was followed in turn by successful appeals in a number of other high profile cases: Maguire Seven, Tottenham Three, Cardiff Three, Stefan Kizsko, Judith Ward, Darvell brothers, Taylor sisters, and others. While each of these successful appeals was the result of specific grounds of appeal, and thus an acceptance by the Court of Appeal in each appeal that there had been an individual miscarriage of justice, press reporting on these cases was quite different. The media sought to link these successful appeals, and use them as evidence of a systemic failure of the English criminal justice system which, in the often repeated newspaper phrase of the time, was now itself 'on trial'.¹⁰ While there was never a consensus as to the nature of this systemic failure, the certainty that the successful appeals were evidence of some general if submerged failures of criminal justice, also led to the belief that there must be large numbers of miscarriages of justice waiting to be identified. The sense that the high-profile successful appeals represented a tip of some quite enormous iceberg was well evident in newspaper articles at the time. With influential journalists making this claim, it was not surprising that others repeated it.

[A] regrettable proportion of those in prison are innocent of the crimes for which they were convicted.¹¹

6 Of course these two questions are closely connected. However we feel that the useful assessment offered by others, such as A. James, N. Taylor and C. Walker, 'The Criminal Cases Review Commission: Economy, Effectiveness and Justice' (2000) Crim LR 140 is more characteristic of the second question, whilst the analysis we offer here is more representative of the first.

7 These self-proclaimed ambitions are set out in summary at the beginning of the Annual Reports: Independent, Thorough, Investigative, Impartial, Open, and Accountable.

8 A statutory duty requires these Annual Reports to be produced, Criminal Appeal Act 1995, Sched 1 para 8.

9 *Royal Commission on Criminal Justice*, Report Cm 2263 (1993) ch 11 para 11.

10 See our analysis of the construction of a crisis of confidence in the criminal justice system by the media in the early 1990s, in R. Nobles and D. Schiff, *Understanding Miscarriages of Justice: Law, the Media, and the Inevitability of Crisis* (Oxford: OUP, 2000) ch 4.

11 The well-known writer and broadcaster Ludovic Kennedy, whose books relating to a number of miscarriages have been so significant. 'If he were guilty, he could be free' *The Independent*, 19 March 1992.

Waiting in the wings of the Court of Appeal is a legion of further, less notorious but equally shocking cases.¹²

Thus, if the task facing the Commission was to be read off from the newspaper articles that preceded its creation, it needed to deal with a potentially enormous number of miscarriages of justice. And this enormous number was a result of systemic failure whose exact nature and the consequences of tackling it, had not been adequately identified or assessed. This statement of the Commission's task points immediately to the suspicion that it is an impossible one. An alternative response, reflected in the work of the Commission and its Annual Reports, is that the first job of the Commission was to seek to identify, through its own practices, a definition of the problem represented by miscarriages. Once identified, then the problem could be 'successfully' addressed.

Assessing the work of the Commission

By reference to what went before

Before April 1997,¹³ convicted persons in England and Wales who had exhausted their rights of appeal looked to the Home Secretary. He, with the aid of the civil servants advising him, could exercise his powers to initiate investigations into their cases. Such investigations would be arranged either within the Home Office, mainly relying on new police enquiries, or by setting up an independent (usually judicial) inquiry. And following such investigations, where the Home Secretary saw fit he could refer cases to the Court of Appeal (Criminal Division)¹⁴ or grant pardons.¹⁵ The reality underlying this opportunity for convicted persons was a small department at the Home Office, known as the CCU (Criminal Cases Unit), formerly C3 Division,¹⁶ comprising a small number of Home Office civil servants¹⁷ who received between 700 and 800 applications per year involving claims of miscarriage.¹⁸ Investigations in relation to those claims resulted in an average of less than 10 appellants' convictions being referred to the Court of Appeal per year, although the years 1990–1992 saw a slight increase.¹⁹ The work

12 David Rose, author of *In the Name of the Law: The Collapse of Criminal Justice* (London: Vintage, 1996), "Murderer" Waits 42 Months for Appeal after Evidence Withdrawn' *The Observer*, 17 May 1992.

13 The CCRC was formally established on 1 January 1997 and started operating on 1 April 1997, by which time existing cases had been transferred to it from the Home Office and Northern Ireland Office.

14 Criminal Appeal Act 1968, s 17(1)(a).

15 Namely a recommendation to exercise the Royal Prerogative of Mercy, in the form of a free pardon, a conditional pardon or remission of sentence.

16 There was a similar division in the Northern Ireland Office. C3 Division was itself part of the Criminal Policy Department of the Home Office.

17 'In 1995 C3 comprised 21 staff at the Home Office; there were 3 officials in Northern Ireland.' N. Taylor and M. Mansfield, 'Post-conviction Procedures' in C. Walker and K. Starmer (eds), *Miscarriages of Justice: A Review of Justice in Error* (London: Blackstone Press, 1999) ch 11, 233, n 20.

18 Home Office Memoranda to the Royal Commission on Criminal Justice (1991) Part IV 'Consideration of Miscarriages of Justice' 20. Giving the example of 1991, the Home Office stated that 'the readily identifiable costs of the current arrangements are £357,000 a year for staff costs for C3 Division' (17). The scale of operation of C3 can be compared with that of the CCRC by comparing this figure for staffing costs with that of the CCRC for 1999/2000 – £2,838,698 (AR3, 36).

19 For exact figures for the period 1989–92, see the *Royal Commission on Criminal Justice*, Report, n 9 above, 181.

of this Division was heavily criticised during the widespread media attention to miscarriages of justice that followed the successful appeals of a number of high-profile (often terrorist) prisoners, listed in the previous section of this article, in the early 1990s.²⁰

C3 Division, with the Home Secretary, was accused of being slow, inefficient, too passive, lacking independence,²¹ reactive rather than proactive,²² and of showing too great a deference to the Court of Appeal.²³ The Royal Commission on Criminal Justice attributed these failures primarily to the Home Secretary and C3 Division's constitutional relationship to the courts. Being parts of the executive, they could not be seen to challenge the independence of the senior judiciary who staff the Court of Appeal.²⁴ Having diagnosed the situation as embodying a constitutional problem, the Royal Commission recommended a constitutional answer:²⁵ the setting up of a new body, independent of both the executive and the judiciary, to take over C3's functions and the Home Secretary's power of referral.²⁶

The bland figures for the three-year period of the Commission's operation up to 31 March 2000 indicate that the Commission is more active than its predecessor: 80 referrals have been made,²⁷ a yearly average of around three times the number of referrals compared to the Home Secretary. Of the 80 referrals made, 35 have been considered by the Court of Appeal, of which 27 have succeeded on appeal.²⁸ Thus, if the Home Secretary/C3 Division's lack of referrals were due to their lack of independence (in constitutional terms), the problem seems to have been addressed. However, while the Commission is clearly more active than the Home Secretary/C3, its lack of independence from the Court of Appeal derives from more than any constitutional inhibitions felt by the executive and has, as we shall see, *necessarily* survived the replacement of the Home Secretary/C3 by the Commission. As a body with no power to declare miscarriage on the basis of its own standards, but only to investigate cases with a view to having miscarriages recognised by the Court of Appeal, its relationship cannot easily be anything other than one of deference.

20 For an analysis of this media attention, see n 10 above, ch 4. In the Second Report by Sir John May on the *Maguire Case* he makes it clear that he believes many of those criticisms to have been misplaced. 'In the light of such constitutional reasoning I think that most of the public criticism that there has been of what with hindsight have been shown to have been miscarriages of justice, and of the alleged failure of the authorities to do anything about them, has been unjustified.' (HC 296 (1992) para 10.8).

21 For a succinct statement of these criticisms see 'Memorandum by Justice', Appendix 3 to the House of Commons Select Committee on Home Affairs, *The Work of the Criminal Cases Review Commission*, HC 106 (1999).

22 See the quotation from Sir John May, set out in the *Royal Commission on Criminal Justice*, Report, n 9 above, 182.

23 The overall significance of C3 Division in remedying miscarriages of justice was thought to be minimal. 'Sir David Napley told the Home Affairs Select Committee, in 1982, that he was not aware of any miscarriage of justice that had been corrected as a result of an initiative by the C3 Department. I do not know if much has changed in the ten years since that remark. (C. Mullin, 'D.N.Pritt Memorial Lecture' (Spring 1992) *Socialist Lawyer* 17) See also R. Pattenden, 'Criticisms of the Home Office' in *English Criminal Appeals 1844-1994* (Oxford: OUP, 1996) 387-396.

24 It should be remembered that section 17 referrals involved cases where previous appeals had already been heard by the Court of Appeal (or leave to appeal had been rejected) and the convictions upheld.

25 *Royal Commission on Criminal Justice*, Report, n 9 above, ch 11, paras 1-11.

26 And similarly, to take over the role of the Northern Ireland Office and Northern Ireland Secretary in this regard.

27 AR3, 18.

28 AR3, 19. Up to the period covered by the third Annual Report, at 31 March 2000, 45 referrals were waiting to be heard by the Court of Appeal.

By reference to the Court of Appeal's standards

While deficiencies of the Home Secretary/C3 Division were attributed to the executive's undue deference to the Court of Appeal, this was not a consequence of any statutory restriction. The Home Secretary could refer cases 'if he thinks fit'. Somewhat surprisingly, given the diagnosis of the Home Secretary/C3's deficiencies, the Commission is expressly *required* to defer to the Court of Appeal's standards and practices. Its principal responsibility as set out in sections 9 and 10 of the Criminal Appeal Act 1995 is to refer convictions to the Court of Appeal. The conditions for making such referrals are set out in section 13(1)(a), namely 'that there is a *real possibility*²⁹ that the conviction, verdict, finding or sentence would not be upheld [by the Court of Appeal] were the reference to be made'. Although there is some controversy over the exact meaning of 'real possibility', what is clearly indicated by section 13 is that the standards to be applied are to be taken from the practices of the Court of Appeal. 'The judgment required by the Commission is a very unusual one, because it *inevitably* involves a prediction of the view which another body (the Court of Appeal) may take.'³⁰ Thus, where the views of the Commission and the Court differ as to what sorts of errors constitute a miscarriage of justice, it should be the Court's views that prevail. On the other hand the Commission does not have to regard a case as a 'racing certainty'³¹ before making a referral. This implies that, provided the kind of error identified by the Commission is of a type recognised by the Court, the Commission does not have to be certain that the Court's conclusion as to its implications for the safety of a verdict will match its own before making a referral.

The 'real possibility' test involves a separation of standards articulated by the Court of Appeal from discretionary judgments as exercised by the Commission, which is a difficult separation for the Commission to operate. It presupposes that the Commission is able to identify the standards of the Court, and thereby understand what it should itself look for.³² In carrying out this task, the Commission has to be careful not to be too conservative in its assessment of the Court's practices, leading to an accusation that it is falling into the unduly deferential role attributed to the Home Secretary and C3 Division. And for those who had hoped that the CCRC would be *more* liberal in its interpretation of miscarriage than the Court of Appeal, even success in this task might still represent undue deference. At the other extreme, should the Commission be seen by the Court of Appeal to take an unduly liberal view of the appropriate standards, it might well face criticism from the Court for wasting its time.

The treatment of two referrals to the Court of Appeal demonstrates the difficulties involved. The referral of *Bentley* led to the re-assessment by the Court of the judge's summing up at the original trial in 1952 by reference to contemporary views of what constitutes acceptable comment. It is not clear from its judgment³³ whether the Court

²⁹ Our emphasis.

³⁰ Lord Bingham CJ in *R v CCRC, ex p Pearson* [2000] 1 Cr App R 141, 150 (our emphasis).

³¹ *Ibid* 149.

³² The decision to refer is taken following case review by a committee of at least three Commissioners, as required by the Criminal Appeal Act 1995, Sched 6, paras 2(a) and 3(a). On the other hand, the decision not to refer may be taken either by such a committee, or by a single Commissioner (AR3, 8). Committees have determined to make 80 referrals and 178 non-referrals over the three years (AR3, 18).

³³ *The Times*, 31 July 1998; [1999] Crim LR 330. In his comment on this case in the Criminal Law Review J. C. Smith recognises its potentially 'alarming implications' in that it anticipates 'rewriting of legal history' and how 'The coming of the Criminal Cases Review Commission has highlighted the problem' (331-332). See Francis Bennion's trenchant criticism of many aspects of the final appeal in this case, 'Rewriting history in the Court of Appeal' (1998) 148 NLJ 1228.

was seeking to establish a general rule that cases can be re-assessed in the light of contemporary standards, or was simply seeking to exercise discretion in the exceptional circumstances of the instant case. Either way, it seemed to indicate that a body whose task was to apply the standards of the Court of Appeal, might decide that a conviction whose safety would be suspect under current practices would be an appropriate case for referral.³⁴ This seems to have been the Commission's approach in their referral of *Gerald*. Convictions in 1987 for conspiracy to cause grievous bodily harm and causing grievous bodily harm were obtained after admissions had been made. However, those admissions were made following restrictions on the defendant's access to a solicitor and without the defendant being given an explanation for the denial of such access. Without legally valid reasons such restrictions would not have been acceptable in the light of more recent developments in case law surrounding section 58 of PACE (the Police and Criminal Evidence Act 1984). In rejecting the appeal, the Court 'expressed surprise' that the Commission had seen fit to refer the case at all.³⁵ The Court did not accept that it was open to the Commission to regard a change in standards as a basis for the exercise of *their* discretion, even though such options were open to the Court of Appeal. In rejecting the appeal the Court stressed that 'there is no new evidence and the evidence and the points which form the substance of the appeal were never canvassed in evidence or argument at trial.'³⁶ The Court expected the Commission to operate on the basis of the Court's general practice, which is to exclude new arguments unless they are linked to the discovery of new evidence, rather than as a response to a change in legal standards or procedures.

What this points to is a restricted task for the Commission. It may take an optimistic view of how the Court might treat new evidence, whether in the form of new facts or of new information about irregularities in the course of the police investigation or trial (the Court's usual and conservative basis for quashing a conviction on appeal). But it may not take an optimistic view of the Court's willingness to depart from this general approach, under its broad general discretion to quash any unsafe conviction on any grounds whatsoever.³⁷ This task is too narrow for many of those, particularly in the media, who take a broad view of what constitutes a miscarriage of justice. In particular, it does not allow the Commission

34 Such a conclusion is supported by the later cases of *Johnson*, *The Times* 21 November 2000, and of *O'Brien* [2000] Crim LR 676. 'Although the substantive criminal law in force in 1988 had to be applied, the conduct of the investigation of the case, the conduct of the trial, the directions to the jury and the reliability of the evidence were to be judged according to present day standards.' However this seems to contradict dicta from the earlier decision of *Ward* (1993) 96 Cr App R 1: 'in order to determine whether there were material irregularities ... we must ... apply as best we can the standards of what was considered to be proper and regular at that time.' (per Glidewell LJ at 23)

35 Transcript of judgment, 3 November 1998, 2; [1999] Crim LR 315, 316. In fact, since no previous appeal or leave to appeal had been made in this case, the Commission had taken the unusual step of referring the case under section 13 (2), where 'there are exceptional circumstances which justify making it.'

36 Transcript *ibid* 2.

37 Criminal Appeal Act 1968, s 2(1) as substituted by Criminal Appeal Act 1995, s 2(1). Despite repeated legislative restatements of the grounds of appeal the Court of Appeal has always had the power to quash any convictions for whatever reasons it deems fit (see D. Schiff and R. Nobles, 'Criminal Appeal Act 1995: The Semantics of Jurisdiction' (1996) 59 MLR 573). In view of this general power to quash convictions (available since the Court's establishment in 1907) discussion of the various grounds of appeal is a somewhat semantic exercise, but one that regularly arises as Court of Appeal decisions seem to impose more restrictive interpretations on this power. See eg A. Clarke, 'Safety or Supervision? The Unified Ground of Appeal and its Consequences in the Law of Abuse of Process and Exclusion of Evidence' [1999] Crim LR 108 (responding principally to the much discussed decision of the Court of Appeal in *Chalkley and Jeffries* [1998] 2 All ER 155; that decision has now been disapproved of in *Togher and Others*, *The Times*, 21 November 2000).

to adopt a sympathetic attitude to defendants (or their lawyers) who have made what, with the benefit of hindsight, are tactical errors in the conduct of their defences.³⁸ Defendants' lawyers will often decide that certain evidence, or lines of argument, may run the risk of undermining their chances of securing an acquittal. Traditionally, the Court of Appeal will not allow such excluded evidence or arguments to form the basis of an appeal.³⁹ This is, despite the rhetorical commitment to the unacceptability of an innocent person's conviction, simply not a practical role for an appeal court, as it seriously threatens the ability of the system of trial to provide a final verdict in a sufficient number of cases.

The Commission's sensitivity to the Court's needs and traditional attitudes to such matters has been noted, and upheld, by the Divisional Court in *Pearson*.⁴⁰ This case involved a defendant who wished to appeal against her conviction for murder on the ground of battered woman's syndrome, even though her original defence at trial had been that she had not been the perpetrator of the attack. The Commission declined to refer her case to the Court of Appeal on the basis that evidence of her mental state (something that would have been presented at trial if she had run the battered woman's defence) would not now be admitted by the Court of Appeal. *Pearson* sought judicial review. The Divisional Court, denying the application, affirmed that the task of the Commission was to predict and make a judgment on the response of the Court of Appeal, a task that involved paying attention to how the Court articulated its standards and practices. In his evidence to the Select Committee on Home Affairs, the Commission's Chairman even accepted that a correct interpretation of the Court of Appeal's standards could give rise to a quantifiable limit to the number of successful referrals that it could expect to produce. If 100 per cent of referrals were successful, it could justifiably be accused of being too cautious. If 90 per cent were being rejected, this could represent evidence of inadequate screening.⁴¹

By reference to a thorough investigation of each case

If the usual standards and practices of the Court of Appeal define the Commission's task, then its principal operative role becomes one of investigation.⁴² The job of investigation is limited. This is because, as we described in the previous

38 The appeal by Anthony Steel as described by the investigative journalist and producer of the television series 'Rough Justice', Peter Hill, is a good example of this situation ('Finding Finality' (1996) 146 NLJ 1552).

39 The traditional test is set out in *Ensor* [1989] 2 All ER 586. It was modified in *Clinton* [1993] 2 All ER 998. The organisation Justice remained critical of the Court of Appeal's attempts to deal with this. See their Report, *Remedying Miscarriages of Justice*, n 5 above, 13. Currently the Court of Appeal must ask itself, simply, whether the verdict is unsafe. But what actions or inaction of lawyers at trial might lead to this conclusion is elusive. For a summary of the 'sort of thing that might succeed' and the 'sort of thing that will not do' see A. Samuels, 'Incompetence or ineptitude of counsel as a ground of appeal' (1997) 77 *The Criminal Lawyer* 1, 3.

40 n 30 above.

41 'It would not worry us if it dropped to 50 per cent but we would be very worried if it was 10.' Sir Frederick Crawford, n 2 above, question 73.

42 The Commission itself is currently made up of 14 members, 10 of whom are part-time. At 31 March 2000 there were 36 Case Review Managers, with new funding which would allow for an increase to 50 in the short term (AR3, 26), and 55-60 in the longer term. (See *The Work of the Criminal Cases Review Commission*, n 21 above, paras 57-58; Government Reply to the First Report from the Home Affairs Committee Session 1998-99, HC 569 (1999), paras 12 and 14.) Under the direction of Commission members it is the Case Review Managers who internally carry out the task of investigation or externally arrange for others, particularly the police but also other agencies, to investigate.

section, it is necessarily linked to the standards for referring cases to the Court of Appeal. For example, if the Commission follows the approach adopted in *Pearson*, it will not seek evidence that could have been, but was not, adduced at the trial. The task of investigation is also limited by the principle that cases should not be referred back to the Court of Appeal simply because one disagrees with an earlier refusal to quash a conviction. (By contrast, those who wish the Commission to identify a basis of miscarriage that would not be accepted by the Court of Appeal would want any bases for an erroneous verdict to be explored.)

While such limitations and restrictions in the standards of what counts as a miscarriage will ordinarily reduce the burden of investigation, this is not always the case. The restrictions imposed by the Court of Appeal make it essential for those who doubt the safety of a conviction to seek new evidence, however weak in itself, so as to allow the Court of Appeal to entertain a fresh appeal. The solicitor Gareth Peirce, in her evidence to the House of Commons Select Committee on Home Affairs, described the difficulties that the Court of Appeal had made for itself when it rejected what she regarded as the strongest possible evidence of the innocence of the Guildford Four: the people who did it admitting to the crime. Thereafter a further appeal was dependent on finding some extra, albeit weaker new evidence of innocence, so as to avoid a new Court of Appeal being simply asked to disagree with the conclusions of the earlier one:

There was the endless hope of trying to find something that was relatively minor, like a missing alibi witness, as compared with the impact of the people who did it. That was always a trawl for second or third best evidence to try to re-open it.⁴³

In cases where the Commission has formed the conclusion, on the basis of the existing file, that a conviction is unsafe, the role of investigation is to find 'something new' in order to justify a referral to the Court of Appeal. Such investigations are designed to overcome the hurdles created by the standards of the Court of Appeal, rather than to assist the process of forming a view on the safety of the original conviction. The Commission has a difficult task in identifying which cases fall into this category, where extra-ordinary investigations are required for strategic reasons, and those which fall into the more routine category, where they are looking for new evidence that will, in and of itself, cast doubt on the safety of a conviction.

The burden of investigation is also linked to the rules surrounding the prosecution's duty to disclose evidence at trial. Following the *Ward* decision, the prosecution were under a routine duty to disclose all evidence (including lines of enquiry that were not followed up) that might have any relevance to a defence case.⁴⁴ In the face of claims that this burden was impracticably high, the Criminal Procedure and Investigations Act 1996 reduced the burden on the prosecution, returning the situation to where the prosecution are only obliged to disclose material that they judge, in the first instance, might 'undermine the case for the prosecution', or in the second instance 'might be reasonably expected to assist the accused's defence as disclosed by the defence statement.'⁴⁵ But what role does this changed practice leave to the Commission? Do they limit their investigation of undisclosed evidence to matters that were covered by the prosecution's duty at trial

⁴³ *The Work of the Criminal Cases Review Commission*, n 21 above, Minutes of Evidence, question 34.

⁴⁴ Namely 'all the material evidence which the prosecution have gathered and from which the prosecution have made their own selection of evidence to be led.' n 34 above, 25.

⁴⁵ ss 3(1)(a) and 7.

or, for the cases being reviewed by them, do they investigate to the higher standard which applied between *Ward* and the coming into force of the 1996 Act?⁴⁶

Even if investigations are limited to matters that create a 'real possibility' of a successful referral,⁴⁷ the process of investigation is potentially open-ended. The Commission can look into the process by which the police and the prosecution generated their case, to see if there has been a procedural irregularity or evidence that has been overlooked. In addition, it can commission new tests, or follow up new lines of enquiry that were not apparent at the time of the original trial. In essence, it is involved in the task of proving a negative proposition – that there is nothing as yet unfound that could throw doubt on the safety of this conviction. The open-ended nature of such enquiries is compounded by the lack of legal training and resources open to prisoners⁴⁸ who claim to have suffered miscarriages of justice.⁴⁹ Some applications amount to little more than an assertion of innocence.⁵⁰

With such an open-ended task, it is the rigour of the investigation that justifies the conclusion that nothing went wrong, or was overlooked.⁵¹ But while the overall nature of the Commission's task (proving the absence of error) leads them to adopt practices that may accurately be labelled as 'meticulous',⁵² there is no objective standard to identify a complete and thorough investigation. Conversely, while the Commission must be seen to be meticulous, there is no minimum standard identifying what level of investigation is deserving of such a label.

Over the three years since its inception, the Commission has altered the manner in which it undertakes this task, in order to make it manageable. From

46 It can be anticipated that with the coming into force of the Human Rights Act 1998 challenges to the application of the Criminal Procedure and Investigations Act 1996 with respect to the right to a fair trial will arise. (On whether such challenges are likely to be successful, see S. Sharpe, 'Article 6 and the Disclosure of Evidence in Criminal Trials' [1999] Crim LR 273.) The Courts will articulate standards in response to those challenges. And then the Commission will have to determine whether the standards adopted address the issue of the safety of convictions. The Act will thus exacerbate the problems that result from determining the extent to which changed standards can be taken to have retrospective effect on the safety of convictions.

47 Which, as the Commission says itself, will (except in exceptional circumstances) generally involve 'looking for arguments not raised, or evidence not adduced, in the proceedings that led to the conviction.' (AR3, 9).

48 There are, however, a few amazing examples of how prisoners have developed extraordinary expertise in order to substantiate their own claims of miscarriage. See, eg K. Callan, *Kevin Callan's Story* (London: Little Brown & Co, 1997); 'Triumph of a wronged man' *The Guardian*, 26 April 1997.

49 Though the number of applications prepared with the assistance of lawyers has risen from 10 per cent to 30 per cent. The Commission has claimed specific responsibility for this since it has made considerable efforts to increase awareness of the availability of legal assistance (AR3, 20). See also the oral evidence of the Commission's Chairman (Sir Frederick Crawford) and Chief Executive (Mrs. Glenys Stacey), n 2 above, questions 50–56.

50 'There is considerable variation in the presentation of new applications to the Commission. At one end of the spectrum, the Commission may receive no more than a letter or application form, with no analysis of the issues for consideration. At the other extreme, the Commission has received submissions from solicitors raising as many as 200 issues, and accompanied by large volumes of documentation.' (AR2, 8.)

51 In its second and third Annual Reports the Commission stresses the importance of this. 'Thoroughness is one of the aspects of prevention, and superficially attractive though it may be to skimp on thoroughness, the costs of mistakes leading to resubmission of cases, judicial reviews and diminished public confidence in the professional competence of the Commission, are likely to be overwhelmingly greater.' (AR2, 7, AR3, 7.)

52 Gareth Peirce used the phrase 'meticulous to a fault' about the CCRC in her oral evidence to the House of Commons Select Committee on Home Affairs on 15 December 1998 (n 21 above, Minutes of Evidence, question 2). Thereafter the designation 'meticulous' has become a common way of referring to the Commission's work.

the very start, it did not review cases at its own initiative,⁵³ but relied on applications from convicted persons. However, and in addition, the Commission has reduced the extent to which its investigators are encouraged to identify and investigate grounds for appeal which are neither raised by the applicants, nor obvious from their applications. The House of Commons Select Committee on Home Affairs, who urged it to target its resources, has supported it in this change of approach.

The key is to identify at an early stage the most critical issues and then to examine those in detail without becoming bogged down or sidetracked by matters which are of less importance.⁵⁴

The manner in which this instruction has been put into practice is quite subtle. The Commission has introduced a new screening stage into its procedures. The initial stages of an investigation remain unchanged. On receipt of an eligible application,⁵⁵ an investigator will assemble all of the documents that would ordinarily go before the Court of Appeal. 'The caseworker will typically study the trial judge's summing up, the court of appeal file, documents preserved under Section 17 of the Act, and unused police and Crown Prosecution Service (CPS) documents and other materials.'⁵⁶ Even if the applicant has identified no substantial issues, these documents will be reviewed before coming to a decision that nothing had gone wrong. A new screening procedure, introduced during the second year of the Commission's operation, has reduced the extent to which its investigators are able to go beyond these documents before reaching a conclusion on the potential for a successful appeal. If an investigator, after 'only a few caseworker-days of effort', can see no grounds for a successful appeal, he or she must present the application to a single Commissioner who must decide whether it ought to be rejected.⁵⁷ This screening procedure has curtailed a potentially fuller investigation of every case. Only those cases that can demonstrate the likelihood of a successful appeal within this allocation of caseworker's time receive a full investigation. The significant savings in caseworkers' time which this procedure has achieved,⁵⁸ is at the expense of the more open-ended enquiries which could previously have been undertaken in connection with every application.⁵⁹ This new procedure allows a decision to reject at this stage to be taken within 30 days of the application.⁶⁰ And it permits substantial savings to be made in the staff resources

53 It has the power to do so under section 14 Criminal Appeal Act 1995, but has not acted on that power. Also, there has been no reference to it from the Home Secretary (s 16) with a view to the exercise of pardon. On the other hand there have been four references to it from the Court of Appeal (s 15) with requests for an investigation and report.

54 n 21 above, para 75.

55 Eligibility primarily refers to the need for applications to have exhausted their rights of appeal (by having appealed, or applied for leave to appeal), unless there are exceptional circumstances (Criminal Appeal Act 1995, s 13(2)). Of the 3,147 applications that had been processed through the eligibility stage over the three years, 24 per cent failed. Sir Frederick Crawford has noted the 'very curious statistic' that despite the increase in legal representation 'the percentage of ineligible cases has not varied very much' (n 2 above, question 53).

56 AR1, 18.

57 Stage 2 Screen, described in AR3 at 14 and 17.

58 We derive this implication about savings from statements made in AR2 and AR3, from the time that the new 'Stage 2 Screen' procedures became operative in May 1999 (AR3, 17).

59 Whereas the possibility of becoming more efficient, and adopting new procedures to that end, was the focus of much of the debate at the Select Committee on Home Affairs on 15 December 1998 (n 21 above), trying to satisfy themselves that efficiency savings had not sacrificed the quality of the Commission's reviews was the focus of the same Committee members' questions to Sir Frederick Crawford and Mrs. Glenys Stacey on 11 April 2000 (n 2 above).

60 AR3, 14.

devoted to checking out the safety of convictions on behalf of persons who cannot come up with grounds of appeal for themselves.⁶¹

The cost of the earlier procedure for investigation, one that allowed the investigators to focus on 'too many' issues, had led to a disquieting state of affairs. While the Commission had always attempted to prioritise more urgent cases, the processing of all residual cases in the order in which they were received had led to a situation in which applications would wait years for decisions to be taken.⁶² Many of these decisions would eventually be taken without further investigation. Delay often created a false impression that investigation was taking place when it was not. All of these factors indicated a need to re-order the allocation of resources (to find a way of giving hopeless ones a quick decision) but it did not indicate a need for more resources. What was more worrying, was the fact that the daily rate of eligible cases needing more than cursory review was exceeding the number of reviews being completed by over a two-to-one ratio.⁶³ The inability to match input and output meant that the time taken to deal with each case would inevitably grow. Even if the queue were re-ordered to dispatch easy cases quickly, the backlog of substantial cases would steadily increase. The Commission asked for extra staff, but ended up receiving less than it asked for, and received these only after a delay. While the House of Commons Select Committee on Home Affairs supported the Commission's need for additional resources, the need for a change in practices was spelled out in the Committee's report.⁶⁴ While the Commission has been wary of pressure to target its resources, arguing that mistakes could result in more work, and a 'loss of public confidence',⁶⁵ it has also played down suggestions that any drop in quality has resulted from their taking a less proactive approach, preferring to treat their earlier more wide-ranging enquiries as a mistake. 'Where we are learning is how to do that thorough job but to do it better by not following false tracks and going off in possibly unprofitable directions.'⁶⁶ As noted in their third Annual Report, the combination of these savings with a drop of 25 per cent in the number of new applications⁶⁷ has led to a situation whereby the resources of the Commission match the workload demanded of it. The daily output of decisions is now sufficient to prevent any further increase in the backlog of cases awaiting decision. The time taken to complete a full review of an application, is currently averaging three years from the date of application, but is not expected to increase any further.⁶⁸

The change in work practices described above and the other 'efficiency' changes that the Commission has undergone makes some sense, both in terms of its ability to reach equilibrium between its workload and its resources, and the changing nature of

61 Further savings in manpower appear to arise from the additional deployment of support staff rather than Case Review Managers or Commissioners to undertake a larger role in the Stage 2 Screen process. See Sir Frederick Crawford's answers to questions 57-60, n 2 above.

62 'If we were to take all of the cases in chronological order that would mean that a very large number of ineligible cases and a very large number that can be dealt with in one to five days would be waiting for several years to be dealt with.' Sir Frederick Crawford, n 2 above, question 40.

63 By the end of the second year of operation 'Although 266 cases were completed, they were replaced by 635 cases deemed eligible for Stage 2 review' (AR2, 12).

64 See text at note 54 above.

65 See n 51 above.

66 Sir Frederick Crawford, n 2 above, question 17.

67 AR3, 17.

68 AR3, 5 and 21. Of course this calculation depends on a number of unpredictable factors. Will the downward rate of applications continue? What are the longer-term implications of some of the short-term priorities that the Commission is applying (in particular giving priority to cases that appear to require fewer casework hours)? Will the funding for the increase in staff requested be forthcoming?

the applications it has to process. At the beginning of its operation, it inherited 284 cases from the Home Office and Northern Ireland office, which included a number of high-profile cases of long standing. These cases are more likely to fit into the pattern of investigation referred to by Gareth Peirce,⁶⁹ where the task of the Commission was to overcome the barrier represented by the Court's practices, ie where new evidence is less the basis of appeal than the pretext by which unsafe convictions can be re-examined with a view to achieving a different result. Some of these cases went back 50 years. The ability to refer such cases to the Court of Appeal, and even achieve successful results, allowed the Commission to demonstrate, right from its inception, that it was more 'successful' than its predecessor was. Having accomplished this, and under pressure to achieve 'effectiveness', 'efficiency' and 'economy'⁷⁰ the Commission could alter the routine basis of investigation to something that was more sustainable on a long-term basis.

By reference to the rhetoric of criminal justice

In the rhetoric of criminal justice, it is not acceptable for an innocent person to be wrongly convicted, or for a guilty person to be convicted through unfair procedures. Within the same rhetoric, justice cannot be priced. Combine the two and one is committed to potentially unlimited expenditure to ensure the safety (both in terms of factual truth and acceptable processes) of each and every conviction. If the Commission followed this approach it would have to decide every doubt in favour of the convicted person: whether to limit itself to the standards of the Court of Appeal or seek to challenge them; whether to search for convincing new evidence of innocence, or evidence that allows the case to be re-heard; whether to check the prosecution's file to see if they carried out their duty, or to re-investigate the investigation to see if any cogent evidence was screened out by the police. Expenditure is only limited by opportunity: justice has only been done when everything possible has been done.

A review body could not hope to give full effect to this rhetoric. It is simply not sustainable on a long-term basis. If restrictions on disclosure and argument apply at trial but not on appeal, the Commission will find itself routinely re-investigating convictions, which could generate a vast increase in the number of applications it receives.⁷¹ In turn, this creates pressure for reforms to the trial process. For example, there is little point in having a restriction on the prosecution's duty of disclosure to the defence if the matters that are excluded will need to be routinely provided to the Commission. Indeed, if wrongful conviction (rather than failure to remedy wrongful conviction) is considered unacceptable the curative practices of the Commission will create pressure to reform trial practices, on the basis that error should be avoided, and not just remedied.⁷²

69 See text to n 43 above.

70 AR3, 7.

71 The correlation between applications and, with information filtering through to potential applicants, increasing numbers of applications is strong. All the evidence suggests that this is a supply-led situation in which, as the chances of success appear to increase for appellants, an increase in the number of applications occurs.

72 The Commission says in its first Annual Report that it will 'mine its growing cases database for features of the criminal justice system conducive to miscarriages of justice that might be susceptible to improvement' (AR1, 7). We are not aware of any example of this to date. The reason is set out in the third Annual Report. 'It is a long-term objective of the Commission to analyse the causes of miscarriages of justice, and to suggest ways in which the criminal justice system might be improved to minimise their occurrence. It would be premature for the Commission to do so on the basis of its 80 referrals to 31 March 2000.' (AR3, 11.)

The Commission is not allowed to take a bold approach to its role. Its statutory authority expressly subordinates it to the Court of Appeal. It can only refer a case where there is a 'real possibility' of the referral succeeding. To quote the Commission's chairman: 'The question is not in the true sense "Is a person in jail who should not be?" The Commission is statutorily obliged to apply a [different] test and that is what it is doing case by case.'⁷³ Nevertheless it may be useful to consider the difficulties of any attempt to adopt a different approach. Indeed the main suggestion of the Select Committee on Home Affairs is that this key statutory provision should be kept under review, with a view to introducing a more proactive mandate, if this should prove necessary.⁷⁴ While there is good reason to expect that the Commission will be found wanting by reference to the rhetoric of criminal justice, just as the Home Office and the Court of Appeal have been in the past, it is difficult to construct a workable relationship between the Commission and the Court of Appeal that does not routinely subordinate the former to the latter, and require both to fail to live up to the full implications of this rhetoric.

By reference to grievance/dispute resolution

At the opposite extreme from the rhetoric of criminal justice, one might see the Commission as a body which has a lot in common with many other investigative and administrative quangos: a body set up for the resolution of grievances and as a forum for disputes. There are a host of tribunals, ombudsmen and other appeal mechanisms that look into the merits of claims by persons with limited resources that they have been treated incorrectly by officials. Whilst finality has value (and saves expense) for all forms of decision making, plausible claims that errors are made in a significant number of cases leads to the establishment of bodies designed to explore this possibility in individual cases. Whereas the tasks undertaken by such bodies are linked to the rhetoric of justice (although not usually criminal justice), such rhetoric is invariably tempered by the resources available to them. As a consequence, the rhetoric of justice is reconstructed into the language of management. For example, in assessing the work of such bodies, delay is seen as an undesirable quality. There is a considerable emphasis on clear-up rates – the number of cases that can be resolved and the time taken to achieve this task.

Many of the features of the Commission fit this pattern. The Commission is, as described above, essentially reactive. It relies on applications by aggrieved persons, and expects those persons to identify possible grounds of appeal. The original organisation of work meant that the Commission, which had twice the caseworker resources of C3 Division,⁷⁵ was spending considerably longer on each case.⁷⁶ This created a situation in which the time taken to resolve applications would steadily have risen to infinity, having already risen from a backlog of 851 cases awaiting review at the end of year one⁷⁷ to one of 1,133 after only two years of operation.⁷⁸ In the face of this situation the Home Office provided more resources and the Commission altered its working practices (as described above,

⁷³ n 2 above, question 73.

⁷⁴ n 21 above, para 24.

⁷⁵ AR1, 7.

⁷⁶ 'The case intake estimate has proved satisfactory, but the Commission has much wider statutory obligations and investigative powers than the CCU; experience has shown that they demand a correspondingly higher level of resources.' (AR1, 7.)

⁷⁷ AR1, 17.

⁷⁸ AR2, 12.

and in other ways) in pursuit of 'effectiveness (achieving high output standards), efficiency (maximising output from its human and physical resources) and economy (minimising the cost of those resources)'.⁷⁹ With the drop in applications in its third year of operation, the input and output rates have been brought into equilibrium.

The current equilibrium leaves the Commission with an average three-year delay⁸⁰ in deciding whether or not to refer the cases that receive a full review (those that show grounds in the original application and that require more than minimal investigation). The Commission has tried to concentrate its resources on those applicants who are in prison. They are also sensitive to ill health and age.⁸¹ It is thus trying to minimise the time spent in prison by the small number of prisoners who will manage to achieve a successful referral prior to their release from prison, and also trying to avoid a situation where a successful referral is only of posthumous benefit. It has also ring-fenced a minimum amount of resources to deal with persons who have completed their sentences (or had non-custodial ones) to ensure that such persons have some prospect of a decision, whatever the delay.

The Commission would like to obtain more resources, and to reduce the period of delay facing applicants. It is particularly concerned that delay will deter meritorious applications.⁸² It may well be correct in this. Persons who protest their innocence are ordinarily ineligible for parole, and there are psychological costs in protesting innocence, and hoping for a successful resolution, rather than accepting what has occurred. The longer the delay, the more likely it is that such costs will deter applications. While delay is a negative quality for the Commission, it may not be so from the point of view of the Home Office, which provides the Commission with its funding.⁸³ The current delays ensure that successful referrals by the Commission are only of most significant benefit to those serving long prison sentences. These are the persons for whom miscarriages of justice represent the strongest grievances. A more efficient service, with shorter delays, increases the benefits to persons with lesser sentences. Increased supply (more resources) creates a risk of increased demand. Delay operates as a crude form of rationing.⁸⁴ However, delay is not simply a function of the work of the Commission, but also that of the Court of Appeal. At present, the Court of Appeal is taking 185 working days to deal with the cases referred to it by the Commission.⁸⁵ If the Commission were able to deal with its backlog more quickly, this would, without more resources being put into the Court of Appeal, translate into a longer backlog in cases awaiting decision by the Court.

The cost of delay to those whose applications are thereby deterred cannot be calculated. But for those who apply, the greatest cost of delay falls upon those persons whose appeals are, eventually, successful.⁸⁶ They end up staying in prison

79 AR2, 6 and AR3, 7.

80 'That accumulation represents nearly three years of casework for the staff in post, which translates into a comparable delay in reviewing some cases.' (AR3, 1.) See also 'Stage 2' at AR3, 21.

81 AR3, 22.

82 '[T]here is a serious risk that meritorious applicants may not consider it worthwhile to apply to the Commission.' (AR3, 26)

83 See *The Work of the Criminal Cases Review Commission*, n 21 above, para 16, in which funding arrangements year by year are set out.

84 For a more sinister example of this, see Peter Hill's article 'Another weapon in the armoury' (1995) 145 NLJ 1620.

85 AR3, 18.

86 Although, ironically, one may also have some sympathy with 'guilty' persons who live in hope of having their convictions quashed, who would come to terms with their situation more quickly with an early refusal to refer their cases to the Court of Appeal.

for the longer period represented by the delay. By providing increased resources to the Commission one might (assuming that this did not increase applications) reduce the time spent by such persons in prison. While the rhetoric of criminal justice makes a day in prison for the wrongly convicted an unacceptable injustice at any price, a more pragmatic approach allows one to look at the costs of avoiding such matters. In the Commission we have a body which, for a budget of £5.5 million (in 1999–2000)⁸⁷ referred 36 cases in that year to the Court of Appeal.⁸⁸ At its current success rate of approximately three-quarters,⁸⁹ this represents, as a crude calculation, an expenditure of approximately £200,000 per successful referral.⁹⁰ Against this, one should perhaps deduct the £24,000 per year saved by removing ‘innocent’ persons from prison.⁹¹ For prisoners who would otherwise spend another eight-and-a-third years in prison, the service is relatively cost neutral. If one adds on the cost of compensation, which is likely to rise with the increased time spent in prison, paid to those wrongly imprisoned, the length of unfinished sentence required to justify the cost of a referral and successful appeal drops even further. While pricing successful referrals may seem iniquitous, the Commission’s Annual Reports invite just such calculations. They provide a set of accounts, and present themselves as a body that is ‘efficient’ and ‘accountable to the public’ for its performance. They present successful appeals as their successes, things that justify their existence.⁹²

By reference to its contribution to enhancing public confidence in the criminal justice system

A grievance resolution approach invites one to concentrate on the number of successful referrals in order to assess the work of the Commission. Resources spent on unsuccessful applications are simply part of the cost of identifying those with genuine grievances that require remedy. By contrast, the Commission’s role in ‘enhancing public confidence in the criminal justice system’⁹³ causes one to place value on both successful and unsuccessful applications. The Commission’s successful referrals demonstrate the increased ability of the criminal justice system to identify and remedy mistakes. On the other hand, with this increased ability, one can have greater confidence that all other cases represent safe convictions, despite the claims made by some prisoners and the publicity they engender. Whilst only 36 cases were referred in 1999–2000, there were 674 decisions on eligible cases not to refer,⁹⁴ on the basis that there was no ‘real possibility’ of success. This represents a certificate of the safety of each of these convictions as compared to the potential lack of safety of the 36 referrals, at an average cost of only approximately £8,000.⁹⁵

The Commission’s role in affirming the safety of convictions may provide some key to the processes by which it has reached some equilibrium in its operations. As

⁸⁷ This figure is the Grant in Aid received by the Commission in the year ending 31 March 2000 (AR3, 30).

⁸⁸ AR3, 18.

⁸⁹ 27 out of 35, which includes sentence appeals (AR3, 18).

⁹⁰ This may be crude but is nevertheless a relatively accurate statistic.

⁹¹ See James, Taylor and Walker, n 6 above, 146.

⁹² This is represented symbolically in the second and third Annual Reports by the way in which the case histories of successful (and the relatively few unsuccessful) appellants referred by the Commission are highlighted in what are described as case cameos.

⁹³ AR3, 6.

⁹⁴ See Table 1, AR3, 16.

⁹⁵ Again, this is obviously a crude statistic that ignores the processing of cases for eligibility as well as the other responsibilities of the Commission.

noted above, delay by appeal bodies in dealing with grievances is seen as unsatisfactory. Delay prompts two possible responses: to increase the post-appeal body's resources or to lower the resources that it is allowed to spend on each case. While the second response is cheaper, it reduces the post-appeal body's ability to provide a convincing reassurance that the majority of the cases that it reviews have been correctly decided at trial and initial appeal. In the Report of the Select Committee on Home Affairs we see an interesting example of these conflicting aims being brokered. The Committee joined with the Commission in accepting that the rising delays indicated a need for increased resources (ie that the functions of the Commission could not be carried out by simply reducing the meticulous way in which they currently operated). But they joined with the Home Office in accepting that the Committee could not continue in making open-ended demands for increased resources.⁹⁶

It is difficult to assess whether the expenditure of the Commission represents 'value for money'⁹⁷ without examining the nature of what it is expected to achieve. Is any 'loss of public confidence' caused by miscarriages of justice something that can be alleviated by the routine if meticulous investigations and referrals undertaken by the Commission? Our research, described at greater length elsewhere,⁹⁸ leads us to the conclusion that the 'loss of public confidence' in the criminal justice system, which led to the creation of the Commission was something generated by, and existing within, the media. Although everyday encounters with police malpractice and wrongful convictions may establish a loss of confidence for members of suspect communities, the manner in which the general public both experience the criminal justice system, and receive information on how to assess its capabilities, is through the media. This does not deny that the media's reporting arose out of high profile cases in which there was evidence, and widespread belief, of malpractice by police and prosecution agencies, and a perceived reluctance by the authorities to respond to these. But the perception that such cases would damage the confidence of the general public, indeed that it would generate a 'crisis' in their confidence, was the consequence of the media's willingness to see these cases as symptomatic of general deficiencies in the criminal justice system. The media's consensus that things were wrong, and that something needed to be done, was not matched by a consensus on what needed to be done. The period, during which the media perceived that confidence in the criminal justice system (and in the safety of the convictions that it produced) was in crisis, was a period in which those proposing radical reforms could obtain access to the media. But a sustained analysis of and consensus on the implications of such reform for the workability and expense of the everyday practices of investigation and trial did not accompany this media talk.

It is in the context of this media perception of 'crisis', that the Commission may have made its greatest contribution to enhancing public confidence, before it carried out a single investigation. Its creation represented the most significant institutional reform to arise out of the media's reporting of high-profile miscarriages of justice. While it offered little reason to believe that miscarriages would not arise, it offered a greater prospect of them being remedied. For the media, to whom 'crisis' is a signal of the need for change, more often than it can ever be a consensus on what that change should be, the promise to create a body

⁹⁶ n 21 above, para 58.

⁹⁷ Or whether the Select Committee's intervention successfully brokered these conflicting aims at the right point.

⁹⁸ See n 10 above.

who could do more than the Home Secretary/C3 Division to remedy miscarriage (due to a different constitutional relationship and more resources) did much to defuse the perception of crisis. From this perspective, the 'success' of the Commission in enhancing confidence was tied up, from its beginning, in its ability to achieve more than its predecessor: to investigate more cases, more fully, and make more referrals.

The curative property of introducing new appeal mechanisms has an old history. The creation of the Court of Criminal Appeal in 1907 had a similar effect. From its inception the Court reduced the sense of crisis, as expressed in the media, that had preceded its establishment. Consider the following quotation taken from a leading article in *The Times* in 1907 at the time when the Criminal Appeal Bill was being debated in Parliament:

Were there no other reason for creating a Court of Appeal, a strong case would be found for it in the demoralizing irregular discussions in Parliament and in the Press of the verdicts of juries ... If there is to be no Court of Appeal, we must expect to see more and more agitations set on foot or carried on by newspapers ... Something has to be done to counteract, as far as possible, the truly alarming ease with which nowadays a criminal who has been tried and found guilty by his countrymen can induce a large number of well-meaning persons to take for granted his innocence.⁹⁹

If one substituted 'Criminal Cases Review Commission' for 'Court of Appeal' and ignored some of the more dated expressions, this leader could have been written in 1995 when the latest Criminal Appeal Bill was being debated in Parliament. In a similar way to the establishment of the Court of Criminal Appeal, the establishment of the Commission has reduced the sense of crisis about miscarriages of justice that had previously built up in the media.

However, the long-term 'success' of the Commission in enhancing public confidence is more difficult to assess. Do the Commission's routine operations reduce the media's ability to perceive criminal justice as a system that makes mistakes but is unable to remedy them? To some extent this is a circular situation. The Commission's ability to project itself as a successful body allows the media to use its operations as an indicator of where mistakes have occurred, and when they have not. But the factor that most threatens this perception of 'success' is the Commission's need to refer cases that meet the standards and practices of the Court of Appeal. When assessing the safety of a conviction, the media can assess all and any evidence on a *de novo* basis. They can give little or no weight to the fact that a jury or previous court of appeal accepted the defendant's guilt, and they are unlikely to refuse to examine an argument or piece of evidence solely for the reason that it could have been, but was not, produced at trial. Nor will they look at the safety of a conviction in terms of the standards operating at the date of trial, if current standards would contribute to a different outcome. These factors can build to expectations being generated in the media to which the Court of Appeal, and a body working with the same standards as the Court of Appeal, is unable to respond.¹⁰⁰

It is very difficult to form any view as to the Commission's ability to head off media frustration with the Court of Appeal, even if one restricts oneself to the narrow question as to whether it will carry out this role better than the Home Secretary/C3 Division. Home Secretaries have responded in the past to high-profile

⁹⁹ 30 July 1907.

¹⁰⁰ A Note by Just Television to the House of Commons Select Committee on Home Affairs sets out the case against the Commission as an adequate investigative body, and the continuing need for investigative journalism to play a role in remedying miscarriages of justice (Appendix 7, n 21 above).

campaigns on behalf of particular prisoners by referring cases that fell well short of the standards of the Court. One of the best examples was that of the Luton Post Office murder case, referred four times to an increasingly exasperated Court of Appeal.¹⁰¹ There was little evidence here of the constitutional deference that led the Royal Commission to recommend that the Home Secretary's powers of referral should be transferred to the Commission. Some practitioners have also suggested that the vulnerability of the Home Office to political pressure (through campaigns etc) could obtain more resources than are provided by the Commission, more quickly, for the kind of intensive investigation required to get around the usual standards of the Court of Appeal.¹⁰²

The Commission as a site of constructive misreading of miscarriage of justice by different systems of communication

Our earlier work on miscarriages has been 'informed' by autopoietic systems theory, as advanced by Niklas Luhmann and Gunther Teubner.¹⁰³ This is not the place to set out their theories, but their approach, which invites one to consider the law and other systems, as systems of communication reflexively communicating back to themselves, provides insights when one examines the Commission's role.¹⁰⁴ What we have argued in our earlier work on miscarriages is that the media systematically misread the legal system, and that the legal system is incapable of responding directly to the media's reporting of its own operations. Miscarriage of justice, within the routine operations of the legal system, requires the identification of something that has gone wrong with the procedures of investigation, trial and appeal. The media systematically misreads law's claim to determine the 'truth' of a person's guilt, as if the legal system had a scientific commitment to that truth as fact. But, while one of the central features of scientific truth is the replicability of results (the same outcome can be produced on another occasion with different personnel) the legal system cannot hope to live up to this standard. Whereas its rhetoric of 'beyond reasonable doubt' points to a scientific standard of truth, the reality (at best) is that a particular jury, after a particular trial, was extremely certain that the defendant had committed the offence. It does not mean that any jury would inevitably have reached that same verdict. Nor does it mean that this jury, if some part of the evidence produced during the trial process were changed, would ultimately reach that verdict. And it certainly does not mean that a group of scientists, or investigative journalists, after an open-ended inquiry unrestricted by legal procedures, would invariably reach the same result.

The Commission occupies a site where the media, the political system, and the legal system use a common expression (miscarriage of justice) whose meaning is established by the operations of each of these separate systems. What it means for the media and within political discourse, is not what it means within the legal system. The Commission is a body established by Parliament and responsible to

101 See L. Kennedy (ed), *Wicked Beyond Belief: The Luton Murder Case* (St Albans: Granada, 1980); B. Woffinden, *Miscarriages of Justice* (London: Hodder & Stoughton, 1987) chs 6 & 7.

102 See *The Work of the Criminal Cases Review Commission*, n 21 above, Minutes of Evidence, question 3.

103 n 10 above, especially the Foreword, 'A collision of discourses' by Gunther Teubner, and ch 1.

104 We explore these insights further, in relation to criminal justice, in R. Nobles and D. Schiff 'Criminal Justice: Autopoietic Insights' in J. Pribán and D. Nelken (eds), *The Consequences of Autopoietic Law Theory* (Aldershot: Ashgate, 2001) ch 9.

the Home Office. It is staffed by persons who are either legally qualified or have 'knowledge or experience of any aspect of the criminal justice system'¹⁰⁵ (which includes scientists, managers, and even, recently appointed, an investigative journalist). It takes very seriously its task of reporting its operations to the media. Such a body does not dissolve the different meanings of miscarriage of justice generated by law, media, politics, and science, but the proximity of such different personnel does increase awareness of such misreading and its implications. For example, Commissioner lawyers who deal on a day-to-day basis with the spectrum of personnel who work at the Commission, can see the potential for normal legal practices to fail scientific tests, or to be scandalised in media reporting or parliamentary debates. Such proximity also makes the task of managing misreading more manageable.

The task of managing disjunction between law's operations and the understanding of those operations within other systems of communication, most notably the media, should not be exaggerated. While law's claims to produce truth have the potential to generate awareness of their systematic inability to deliver anything close to scientific facts, the routine relationship between law and the media is one in which the media accept convictions as proof of an individual's guilt. Remedied miscarriages of justice provide a regular source of a particular kind of news: individual human tragedy. But the media rarely unite such stories into a general theme of systemic failure. Indeed, such a theme of systemic failure, taken seriously, threatens the media's ability to continue with their usual relationship to the legal system, which is as a cheap and accessible source of stories about violence, greed, evil, etc. Such stories depend on a conviction representing the truth of a person's guilt. This, rather than any view of miscarriage of justice as evidence of the systemic failure of the legal system to separate the guilty from the innocent, is the normal relationship (structural coupling) between the legal system and the media. And this is a relationship which the Commission, with its seal of approval on most convictions which it investigates, helps to maintain.

Conclusion

The success of the Commission cannot be based on its ability to identify and remedy all those cases in which a person convicted was factually innocent or convicted unfairly, or sentenced to an inappropriate term of imprisonment. These matters are unknowable, and can never give rise to statistics against which to compare the work of the Commission. Its success in maintaining public confidence is based on trust, or more particularly, on the media's perception (and those like MPs who might otherwise raise miscarriage cases with the media) that it is a body to be trusted with getting these matters right.¹⁰⁶ Indeed, to the extent that the media believes that the Commission is capable, with enough resources, of identifying miscarriages, the Commission can use the threat of future failure to obtain media support to resist cuts to its budget. In terms of the rhetoric of justice, the creation of the Commission represents a success, not because such rhetoric can be achieved,

¹⁰⁵ Criminal Appeal Act 1995, s 8(6).

¹⁰⁶ There is clear evidence that the Commission understands the need to respond to the media, and to political pressure (see AR2, 19). The Chairman of the Commission seems to reply personally to interventions by MPs: replying to 217 case-related letters from MPs in 1997–98, 231 in 1998–99 and 162 in 1999–2000 (AR1, 10, AR2, 19, AR3, 27).

but because the Commission can do more, and thereby come closer to this rhetoric, than what went before.¹⁰⁷ To date, it has managed the gap between this rhetoric and its own operations. It has established equilibrium between the demand for its services and what the Home Office feels able to afford.

The creation of the Commission has altered the dynamics of post-appeal practices; reviewing applications is no longer the work of a small department buried in the Home Office.¹⁰⁸ The Commission, a body with its own Chairman, Chief Executive, Director of Finance, Commission Members, Case Review Managers, Legal Advisers, and other executive and support staff, has a far greater stake in presenting itself as a successful body than its predecessors did. While it may not be independent of the Court of Appeal, the creation of a body with a separate corporate existence makes it unlikely that post-appeal processes will ever return to the small, passive operation undertaken by the Home Office.

107 Assuming that the introduction and expansion of appeal bodies does not (as it does in other areas such as immigration) allow for a deterioration in the quality of decisions made at a lower level in the system.

108 As Kate Akster of Justice suggested in her oral evidence to the House of Commons Select Committee on Home Affairs: 'I think we take the view that the CCRC is light years ahead of the Home Office in all sorts of ways' (n 21 above, Minutes of Evidence, question 9). In the light of this view, Justice is now less likely to play such a significant role as it did in the past in receiving, reviewing, investigating and campaigning on behalf of alleged miscarriages of justice.

CASES

Forensic Immunity Post-*Osman*

*Rosalind English**

Barristers and solicitor advocates are now liable in negligence for their conduct of a case in court. Not a very radical proposition. But it took four days of debate before a specially convened panel of seven Law Lords to establish that the rule in *Rondel v Worsley* was no longer relevant to modern legal conditions. Forensic immunity is dead. May it rest in peace.

In *Hall v Simons*¹ the House of Lords concluded that the administration of justice would not be brought into disrepute by the threat of litigation against advocates. Courts have greater powers than ever before to stop vexatious claims in their tracks, through their powers of strike out² and their own inherent jurisdiction to prevent relitigation that would amount to abuse of the process of the court.³ The ruling was only unanimous to the extent of abolishing forensic immunity for civil trials. Lords Browne Wilkinson, Hoffman, Steyn and Millett thought that immunity could go altogether. Lords Hutton, Hobhouse and Hope took the view that defence lawyers still need a core immunity in respect of their conduct of criminal proceedings.

In the event, the Lords based their judgment on relatively narrow grounds – the evaporation of most of the justifications for forensic immunity since *Rondel* was decided, and the scope of *Hunter* – but the arguments debated in the course of the hearing covered a range of issues, from the compatibility of the *Caparo* test for the duty of care with Article 6 ECHR, to the continued existence of the cab-rank rule and the precise nature of the duty owed by the advocate to the court and his client. What emerged from the hearing, although the judgments are silent on this, were the first signs of conflict between the House of Lords and the Court of Human Rights in Strasbourg as to the proper application of Convention rights to principles of domestic law.

First, it is worth considering briefly the arguments and the judgment itself.

Forensic immunity

The rule of immunity is encapsulated in the landmark case of *Rondel v Worsley*,⁴ which confirmed that forensic immunity was required for the following:

- (1) to prevent relitigation, otherwise than on appeal, of issues already concluded adversely to the plaintiff by court decision;

* Barrister.

1 [2000] 3 WLR 543. The author gave comparative law assistance to counsel for the respondent.

2 Under CPR r 24.2, the plaintiff must have 'a real prospect of succeeding on the claim.'

3 Set out in *Hunter v Chief Constable of West Midlands Police* [1982] AC 529.

4 [1969] 1 AC 191.

- (2) to ensure that advocates perform their duty to the court as fearlessly and independently as possible, unimpeded by apprehension of actions against them by their clients;
- (3) to allow advocates the general immunity from civil liability which attaches to all persons who participate in proceedings before a court of law.

Some ten years later it became evident that forensic immunity could not simply stop at the court door. In *Saif Ali*⁵ the House of Lords sought to establish what work outside court should also attract immunity. Thus the immunity was said only to attach to particular work which was 'so intimately connected with the conduct of the case in court that it can fairly be said to be a preliminary decision affecting the way that cause is to be conducted when it comes to a hearing.'⁶ Then the whole set of conditions for immunity had to be overhauled when solicitors were granted rights of audience in 1990.⁷

So, forensic immunity, which started life as an exclusive exemption for advocates' acts done in court, was extended, by logic, to an advocate's acts of the same character, done outside the court. This was extended, again by logic, to a solicitor advocate's acts of the same class done in or out of court. It was argued by the respondent clients in *Hall v Simons* that that if this was logic, logic had created a nonsensical situation.

The *Hall v Simons* hearing before the Lords was a culmination of a series of test cases concerning solicitors. The most representative case involved a matrimonial finance respondent, who had been advised by his solicitors to offer his former wife a lump sum from the sale of the matrimonial property. As it happened, the sum was out of proportion with the market value of the house and the husband sought to claim damages against the solicitors in negligence. The solicitors relied on forensic immunity, pointing to the fact that the terms of the settlement were incorporated into a matrimonial consent order, which by statute required approval by the court. The initial question before the Court of Appeal was whether consent orders, drafted by solicitors and approved by the court, were so 'intimately connected' in the *Saif Ali* sense that a subsequent action alleging negligent conduct by the lawyer would be an abuse of the court's process by challenging the earlier judgment. By the time the test cases came before the House of Lords, the respondents argued that they should undertake a root-and-branch review of advocates' immunity altogether.

Let us consider first what the Lords had to say about the public policy justifications for immunity.

Relitigation

This was one of the strongest arguments in favour of preserving forensic immunity. The law should discourage relitigation of the same issues except by means of appeal. As Lord Hoffman said (in a manner refreshingly oblivious to the plain-speaking spirit of the Woolf reforms), 'The Latin maxims often quoted are *nemo*

5 *Saif Ali v Sydney Mitchell and Co* [1980] AC 198.

6 A test based on the 'door of the court' laid down in a decision of the New Zealand Court of Appeal, *Rees v Sinclair* [1973] 1 NZLR 236.

7 Section 62 of the Courts and Legal Services Act 1990 provides that

'A person who is not a barrister but lawfully provides legal services in relation to any proceedings, shall have the same immunity from liability for negligence in respect of his acts or omissions as he would have if he were a barrister lawfully performing those services.'

debet bis vexari pro una et eadem causa and *interest rei publicae ut finis sit litium*.⁸

A person should not be troubled twice for the same reason; and there is a general public interest in the same issue not being litigated over again. Both are principles worth upholding. But is such a broad exclusion of liability for the legal profession necessary to preserve these principles? The majority thought not; in Lord Hoffmann's words, barring all claims in respect of negligent conduct of cases in court in order to prevent relitigation is rather like burning down the house to roast the pig. The first principle – '*nemo debet*' – is adequately preserved by the rules of *res judicata*,⁹ issue estoppel,¹⁰ and abuse of the process of court – in so far as it applies to civil trials.

The second principle – '*finis sit litium*' – is upheld by the broad rule of law precluding collateral attacks on judgments of the courts. This rule against abuse of process, set out in *Hunter*, started life in the criminal courts, and there are obvious reasons why criminal trials should not be re-enacted. Apart from the evidential difficulties of rerunning the criminal proceedings in order to determine whether or not defence counsel had been negligent, there is the difficulty of deciding what consequence should follow should the convicted client show loss of a chance of acquittal.

Nevertheless the majority felt that none of these difficulties were so great that they justified the total ban on actions in negligence by the client, even in a criminal context. It was pointed out that evidential difficulties have not led to an immunity in respect of actions taken by legal advisers outside the court door. The burden of proof is on the claimant, and if his case has become so weak over the passage of time that he cannot prove that the lawyer's negligence caused him loss, then it will be struck out. Where on the other hand there is no difficulty about establishing the link between the lawyer's negligence and the client's loss, there is no reason for giving lawyers immunity from suit.

The other problem with the relitigation argument is determining in what circumstances a subsequent negligence action does in fact amount to collateral attack. Litigation involves a variety of hearings and stages where the involvement of the court moves from the barely perceptible to all-embracing. As Lord Hutton said: 'There is not a single moment of confrontation. The exercise of advocacy extends over a series of processes of which the trial is only one and the advocacy may be conducted as much in writing as orally.'

In these circumstances, the *Saif Ali* 'intimate connection' test is almost impossible to apply. The legal uncertainty created by this condition was one of the many factors which weighed in the balance against forensic immunity.

The dual loyalty argument

This was the next public policy justification due for scrutiny: the duty owed by advocates to the court would be undermined if they were constantly looking over their shoulder at the client in fear of potential negligence suits against them. It

8 572 D-E.

9 *Res judicata* means that even if a decision is based upon a view of the law which is subsequently expressly overruled by a higher court, the judgment itself cannot be set aside: *In re Waring* (No. 2) [1948] Ch 221.

10 An issue estoppel created by earlier litigation is binding subject to narrow exceptions: *Arnold v National Westminster Bank Plc* [1991] 2 AC 93.

would, so the argument went, lead to prolixity and forensic overkill. The advocate would leave no stone unturned in his efforts to protect himself from future negligence actions, thus privileging his duty to his client over his duty to the court.

The difficulty with this claim is that there is little empirical evidence to suggest that advocates would be inhibited in their freedom of speech in court by potential exposure to negligence actions. The Lords accepted that the very real prospect of wasted costs orders¹¹ has not apparently had such an effect (although Lord Hope did not think that any useful analogies could be drawn from the effect of wasted costs orders in the criminal courts).¹²

In any event, under the Access to Justice Act and the new CPR rules, there is far more pressure on those conducting litigation to be economical with their arguments. Today's judge has a 'battery of powers'¹³ to intervene where over-zealous speechifying and discovery threatens to derail a trial.

The cab-rank rule

Both the appellants and the Bar Council argued that losing the immunity would subvert the cab-rank rule. If an advocate is under a professional obligation not to turn away unpromising clients and cases, he should be protected from exposure to liability when things go wrong with that unpromising action. The main problem with this argument is the enfranchisement of a new breed of advocates under the Courts and Legal Services Act 1990, giving rights of audiences to solicitors. The cab-rank rule does not apply to them. If this means that solicitors must be immune when exercising the same function in relation to court proceedings as advocates are immune, how can the cab-rank rule, which does not apply to them, justify their immunity?

In the event, none of the Law Lords attached much significance to the cab-rank rule. It served its purpose in the lexicon of professional ethics, but it could not justify depriving all clients of a remedy for negligence causing them real loss.

Witness immunity

Three members of the House of Lords in the past have treated forensic immunity as closely analogous to witness immunity; Lord Reid in *Rondel* at 230A, Lord Upjohn in *Rondel* at 283C and Lord Diplock in *Saif Ali* at 222A. The rule of witness immunity is that no one may be sued in libel, malicious falsehood or conspiracy to give false evidence, for anything said in court, or in preparation for giving evidence in court.

In this case however the majority felt that this was a false analogy.¹⁴ When the witness is in the box, he owes no duty to his client. Without the rule of immunity, witnesses would be more reluctant to assist the court.¹⁵ Neither of these propositions apply to lawyers. The advocate always owes a common law duty of care to

11 Under s 4 of the Courts and Legal Services Act 1990 the courts may order any legal representative to pay costs wasted by any party as a result of any 'improper, unreasonable or negligent act or omission' on their part.

12 588 G-H.

13 Lord Hoffmann's words at 565 A.

14 In the dissenting speeches it is suggested that a defence lawyer deprived of his immunity might be unfairly exposed to liability in a subsequent negligence action where other participants in the criminal trial, such as the jurors and witnesses, might be equally at fault but immune.

15 A rationale that goes back nearly five centuries: *Cutler v Dixon* 4 Co Rep. 14b, 76 ER 886. The trial process would be inhibited if police witnesses and potential police witnesses were to be exposed to action for things said or done by them during the preparatory stage when they are undergoing this preliminary examination.

his client wherever and whatever he does within the scope of his instructions. The notion that exposure to liability might discourage proper advocacy is intuitive only and has no basis in empirical evidence.

The analogy, if it exists at all, is with expert witnesses, who are something of a hybrid category. A conflict may exist between necessary freedom of speech as witness – especially where the witness is appointed at the court's direction and professional duty to the party appointing him. But this conflict is resolved to an extent by the rule laid down in the CPR that where the expert is engaged on providing expert evidence for use in court his relationship to the court becomes paramount and he enjoys civil immunity attributable to that function.¹⁶ Where necessary in the public interest, the former prevails.

Implications of *Hall v Simons*

So much for the speeches themselves. The abiding interest in the judgment lies not so much in its *ratio* and result – no one expected that barristers' immunity was going to last long into the twenty-first century – but in the way the House dealt with some of the material put before it, particularly that emanating from the Strasbourg Court, and the chain of events set off by the inevitable argument over the decision of that Court in *Osman v United Kingdom*.¹⁷

That case involved a challenge to the public policy immunity of the police against negligence actions taken in respect of the prevention and prosecution of crime. The applicant was a schoolboy who had been the subject of the obsessive attentions of a former teacher at his school. The social services and the police were notified of the teacher's behaviour but failed to prevent the final attack on the applicant's family which resulted in the applicant's father being murdered and serious injuries to the applicant himself. The family's negligence action was struck out on the basis of the rule of public policy immunity set out in *Hill v Chief Constable of South Yorkshire*,¹⁸ and the applicant claimed in Strasbourg that this was a breach of his right of access to court under Article 6. His complaint was upheld by the Court, with significant consequences for the rule of immunity in many other areas of English law.

Although the judgments in *Hall* do not make much of *Osman* itself, there was a broad challenge to the very existence of forensic immunity, based on Article 6 of the ECHR. Long before *Osman* was decided it was well established¹⁹ that the right to a fair trial implies a guarantee of access to court. On the one hand, national legal systems may not put a whole range of civil rights out of reach of its citizens via procedural limitations. On the other hand the Strasbourg Court has repeatedly observed that Article 6 may not be relied upon to generate legal rights that are not recognised in national legal systems. If there is a statutory immunity that prevents potential litigation, that immunity effectively annihilates the civil right – there is nothing on which Article 6 may bite.²⁰

16 See CPR r 35.2 and *Stanton v Callaghan* [2000] 1 QB 765.

17 [1999] 29 EHRR 245. None of the judgments reflect the intensity of the debate that went on during the hearing over what *Osman* means; but see Lord Hutton, para 57.

18 *Hill v Chief Constable for West Yorkshire* [1989] AC 53.

19 *Golder v United Kingdom* (1975) 1 EHRR 524; *Aurey v Ireland* (1979) 2 EHRR 305; *Fayed v United Kingdom* (1994) 18 EHRR 393; *Powell v Rayner v United Kingdom* (1990) 12 EHRR 355 and *Osman v United Kingdom* (2000) 21 EHRR 245.

20 *Ashingdane v United Kingdom* (1985) 7 EHRR 528 – bar on negligence actions initiated by Mental Health Act patients; *Dyer v UK* (1984) 39 DR 246 – s 10 of the Crown Proceedings Act preventing certain actions in tort against the armed forces; *Fayed v UK* n 29 above – immunity from liability for defamation of inspectors appointed under the Companies Acts and *Waite and Kennedy v Germany* (1999) Application No 26083/94 – immunity of international legal systems.

The appellants argued that the rule of forensic immunity was akin to a statutory immunity, and therefore there was no residual duty to which Article 6 could attach. The respondents pointed to section 62 of the Courts and Legal Services which actually describes the bar on negligence actions against advocates as an immunity. This is inconsistent with the proposition that there is no duty at all. The importance of this distinction lies of course with the Strasbourg Court's ruling in *Osman* that an applicant has a right under Article 6 to go to court to argue his case on the merits, if the only obstacle in the way of asserting a duty of care is a broad public policy immunity enjoyed by the alleged tortfeasor.

Osman was a controversial decision which has inspired a whole industry of comment and meta-comment²¹ on a scale rivalled only by the *Factortame* litigation²² and it is not the author's intention to contribute to that production line here. However, some anecdotal material from the hearing in *Hall v Simons* is worth a mention. Some of the more critical commentary on *Osman* comes near to suggesting that somehow the Strasbourg Court has misunderstood the way in which the principles of the English law of negligence are applied. Although the House of Lords did not venture any direct criticism of the result in that case, they did comment that it was incorrect to assume, as the Strasbourg judges appear to have assumed, that the English courts' position on liability was at large; that there was – in the case of *Osman*, at any rate – a pre-existing right for the public to sue the police, and Article 6 attaches itself to that right.

Since their Lordships seemed to be inviting further submissions on precisely what was argued in the *Osman* case, counsel representing the various parties rooted out various memoranda and other official paperwork relating to the proceedings, to support their rival interpretations. This excavation proved highly profitable, but not in the way that anyone could have predicted. One of the documents that came to light was the Resolution of the Committee of Ministers²³ following the judgment in *Osman*. In it, the Foreign Office declared that a circular had been sent to all civil servants and Chief Officers of Police advising them not to claim immunity, because the Strasbourg Court had just ruled that this was contrary to the Convention.

This looked very much like a usurpation by the executive of the judiciary's role in determining the implications of the judgment of another court.²⁴ That this should have happened at all is hardly surprising, because the government is under an obligation in international law to comply with the rulings of the Strasbourg Court. If the judiciary disagrees with one of that Court's findings, what is the government to do? Since their Lordships opted not to refer to this matter in the judgments, there is still no authoritative answer to this question. The Human Rights Act 1998 itself provides no solution, because it only obliges local courts to take Strasbourg's interpretation of individual rights 'into account'. There is no

21 See in particular, L. Hoyano, 'Policing Flawed Police Investigations' (1999) 62 MLR 912; Lord Hoffmann (1999) 62 MLR 159; T. Weir, 'Down Hill all the Way?' (1999) 58 CLJ 4; Lord Browne Wilkinson in *Barrett v Enfield* [1999] 3 WLR 79, and Lord Cooke in *Darker (Personal Representative of David Stanley Docker (Deceased) and Others (A.P.) (Formerly Head and Others (A.P.)) v Chief Constable of The West Midlands Police* [2000] 3 WLR 747.

22 Joined Cases *Brasserie du Pêcheur and Factortame* (C-46 and 48/93) [1996] ECR I-1029

23 Resolution DH (99) 720 'Concerning the Judgment of the European Court of Human Rights of 28 October 1998 in the case of *Osman* against the United Kingdom'. The Committee of Ministers is essentially an enforcement body, which requests information from governments concerning the measures they have taken to implement the Court's judgments.

24 See 'The law lords are already fighting Strasbourg over the Human Rights Act' *The Times*, 29 June 2000.

hierarchy for the national appellate courts *vis-à-vis* the Strasbourg Court; indeed, it has been confirmed recently that no-one is to labour under the illusion that the Strasbourg Court is acting as an appellate body.²⁵

It is unfortunate that the Lords did not take the opportunity in their final judgment in *Hall v Simons* to provide some principled guidance on this important constitutional issue. On the contrary, *Osman* has apparently become such a hot potato that there is increasing reluctance to mention it at all. The lack of discussion devoted to *Osman* in *Hall v Simons* itself means that we still do not know whether their Lordships genuinely believe that forensic immunity was different in essence from the public policy immunity of the police, and therefore *Osman*-proof, since it is a legitimate procedural rule only, or whether they think that *Osman* was wrongly decided, and therefore should not be applied to other areas in the law of negligence.

Whether or not *Osman* was correctly decided, anyone with a passing interest in House of Lords judicial business in the past eighteen months will have noticed one thing: a wholesale retreat from public policy immunity, whether expressed in broad *X v Bedfordshire*²⁶ terms, or in more specific terms, as in the *Hill* case.²⁷ In addition to *Hall v Simons*, the following decisions are a roll call of casualties of this retreat: *Barrett v Enfield*,²⁸ *W v Essex County Council*,²⁹ *Waters v Commissioner of Police for the Metropolis*,³⁰ *Phelps and others v Hillingdon Borough Council*,³¹ *Darker and others v Chief Constable of West Midlands Police*.³²

In each of these cases the House of Lords made reassuring noises that the removal of large parts of public policy immunity would be accompanied by the greater striking-out power of the courts of hopeless claims under the Civil Procedure Rules. Another familiar mantra in these cases is that any floodgates argument could be met by the fact that it is very difficult for the claimant to show that the defendant fell short of the professional standard of care set by the *Bolam* test.³³ That may well be the case, but there is nothing in the *Bolam* test to prevent a 'flood' of claims being taken; the defendant will have to wait for a full hearing on the merits before the burden of litigation is removed.

Certainly if this sudden appellate reluctance to maintain immunity from suit of certain public enterprises is something to do with Article 6, then there is no reason to suppose that junior judges will be any readier to stand their ground against the march of the indignant citizen litigant bearing the ECHR banner,³⁴ particularly at an interlocutory strike-out hearing.

It was in fact the very difficulty of determining complex issues such as justiciability during interlocutory applications or hearings of a preliminary issue that attracted criticism of these kind of strike-outs in the first place. The court deals with hypothetical facts on the statement of claim and statements of principle are formulated which may not represent the true picture.

25 *Locabail v Waldorf Investment Corporation (No 4)* Chancery Division, May 25 2000.

26 [1995] 2 AC 633.

27 *Hill v Chief Constable for West Yorkshire* [1989] AC 53.

28 [1999] 3 WLR 79.

29 [2000] 2 WLR 601.

30 [2000] 1 WLR 1607.

31 [2000] 3 WLR 776.

32 [2000] 3 WLR 747.

33 *Bolam v Friern Hospital Management Committee* [1957] 1 WLR 582, as affirmed in *Bolitho v City & Hackney Health Authority* [1998] AC 232.

34 And indeed positively encouraged to do so by a series of government-sponsored advertisements educating the public about the Human Rights Act during the week it came into force.

This trend started back in June 1999, when judgment in *Barret v Enfield* was handed down. Here the House of Lords reinstated the appellant's action in negligence against a local authority for alleged psychiatric injuries while he was in their care. The claim had been struck out on the basis that a local authority acting within its statutory powers could not be said to have been acting negligently; even if that were not the case, it would not be fair just or reasonable to impose a duty of care in these circumstances. But the House of Lords concluded that justiciability could not be adjudicated upon without a full hearing of the merits. This, observed Lord Browne Wilkinson, demonstrated 'the extreme care which must be taken in striking out claims in this confused and developing area of the law'.

Because *Osman* and its progeny were relative newcomers to the stage at the time, and the passing into force of the Human Rights Act a comfortably long way off, rather more of the discussion in this judgment is given over to the implications of that case. Lord Browne Wilkinson observed that the interpretation laid upon the *Caparo* test by the Strasbourg Court in *Osman* would have the consequence that in any case where it was held that it was not 'fair just and reasonable' to impose a liability, and a claim was struck out, the claimant would be able to go to Strasbourg and argue that he had been deprived of his right of access to court. In other words, Strasbourg would say that the national courts had deprived the claimant of his right to have the balance struck between the hardship suffered by him and the damage to be done to the public interest in the present case if an order were to be made against the defendant council.

Then came *Phelps*, an action for negligent failure of an educational psychologist employed by a local authority to identify that the plaintiff was dyslexic. The original claim had been struck out on *X v Bedfordshire* grounds, but the Lords reinstated the action. A priori decisions about whether it is fair just and reasonable for a body to be liable are misguided, suggested the Lords; it is for the defendant to argue in a full hearing of the merits that the third *Caparo* test should not apply to them.

This time, however, only a glancing reference was made to *Osman*³⁵ and subsequent Strasbourg case law on this particular issue³⁶ and the tenor of the judgments suggest that their Lordships would have reached the decision they did without assistance from their counterparts in Strasbourg.

The *Darker* judgment, handed down on the same day as *Phelps*, removed the broad immunity for things done by the police while they are engaged in the investigation of crime and preparing a case for trial. Here the Lords reinstated an action which had been struck out on the basis that an absolute immunity applied to potential witnesses in a trial.

In *Waters* it was held that the Commissioner of Police did owe a duty of care to police officers under his direction to protect them from harassment and assault by other police officers. The strike-out had been upheld by the Court of Appeal on the basis that there were compelling reasons of policy why bodies in the position of the Police Commissioner should not be made liable for protecting constables under his control from bullying by their colleagues, and therefore the third stage of the *Caparo* test for negligence was not satisfied. Again, the objection to this strike-out was that it was targeted at 'an area where the law is developing'.³⁷ This phrase, it seems, is code for 'vulnerable to Article 6 attack'.

35 Lord Clyde at 808–809.

36 *Z v United Kingdom* Application No 29392/95 where the Commission ruled that the problems attendant on attaching liability to public authorities carrying out complex multidisciplinary functions were not so great as to exclude liability for negligence.

37 Lord Slynn, at 1614 A.

Nevertheless there is nothing in these judgments to suggest that these steps have been taken in anticipation of Article 6, leaving the boundaries of *Osman* in an alarming state of uncertainty. Might witness immunity, for example, be considered different in essence from the public policy immunity of the police, and would it therefore survive an *Osman* challenge? Are there any legitimate limitations on negligence left? Like a free-wheeling loose canon, to coin a popular metaphor, this decision by the Strasbourg Court is laying waste whole areas of English policy, and yet nobody seems to be willing to admit that it is out there.

In any event, no one will deny that the trajectory of these judgments is towards more litigation, with their Lordships firmly set against the assertion of local immunities where negligence is involved. If the Strasbourg Court reaches an adverse judgment in the cases of *Z v United Kingdom* and *TP and KM v United Kingdom*³⁸ the government will be able to tell the Committee of Ministers that the climate has genuinely changed. Whether this is a good thing is a matter for debate, which will no doubt rage on in the national and academic press as the Human Rights Act begins to show its true colours. Perhaps a more interesting question is whether in the light of increased liability, our courts are going to be less generous in determining the quantum of damages.³⁹

38 *TP and KM v United Kingdom* Application No 28945/95 26 May 1998.

39 A partial answer will no doubt be provided by the Lords when they reach judgment in the exemplary damages case, *Kuddus v Chief Constable of Leicestershire* (10 February 2000, unreported).

REVIEW ARTICLE

Sexual Autonomy and Law

Mary Childs*

Stephen Schulhofer, Unwanted Sex: The Culture of Intimidation and the Failure of Law, Cambridge, Mass. and London: Harvard University Press, 1998, xii + 318 pp, hb £17.50.

One striking development of Anglo-American law in the late twentieth century was the intensification of debate regarding legal regulation of sexual conduct. In England, as in other jurisdictions, there was considerable public discussion about the scope and application of laws addressing rape, pornography, prostitution, sexual assault, child sexual abuse and sexual harassment. Some of this debate was instigated by feminists who challenged the split between public and private realms, and who argued that sexual exploitation and coercion were just as much matters for legal regulation as were the presence of coercion or undue influence in matters relating to property rights. The increasing presence of women in formerly male environments has given rise to a greater awareness of the problematic overlap of professional and sexual relationships, and greater awareness of child sexual abuse has contributed to increased concern about the dangers of power misused for sexual purposes. Numerous social phenomena have given rise to new concerns and new calls for legal responses: fear of AIDS, advances in reproductive and contraceptive technology and the increasing political activism of sexual minorities. Sexual matters are now of public concern and debate, and the parameters of the interface of law and sexuality are highly contested. In England, recent years have seen changes in the law of rape, high-profile prosecutions for sexual offences relating to children, registration of sex offenders, legal battles over the rights of transsexuals and same-sex couples, and the defeat and reintroduction of a Bill which would have criminalised sexual relations between teenagers and adult carers, while equalising the ages of consent for same-sex and heterosexual sex.

Most recently, the Home Office has published 'Setting the Boundaries',¹ the lengthy and detailed report of a review of sex offences announced in January 1999. An appendix to that document contains a Law Commission Report entitled 'Consent in Sex Offences'.² This draws upon and adds to the previous work of the Commission on consent in the criminal law, although unfortunately it largely overlooks the growing body of jurisprudential and philosophical writing on sexual behaviour and the role of consent.³

* School of Law, University of Manchester.

1 Available on the Home Office website at: http://www.homeoffice.gov.uk/atoz/sex_offence.htm

2 Available at: <http://www.lawcom.gov.uk>

3 See, among others, D. Dripps, 'Beyond Rape: An Essay on the Difference between the Presence of Force and the Absence of Consent' (1992) 92 Columbia LR 1780; M. Chamallas, 'Consent, Equality and the Legal Control of Sexual Conduct' (1988) 61 S Cal LR 777; N. Lacey 'Unspeaking Subjects, Impossible Rights: Sexuality, Integrity and Criminal Law' (1998) 11 *Canadian Journal of Law and Jurisprudence* 47; N. Naffine, 'Possession: Erotic Love in the Law of Rape' (1994) 57 MLR 10; and the two special issues on 'Sex and Consent' (1996) 2 *Legal Theory*.

Similar issues have arisen in other jurisdictions, including the United States. *Unwanted Sex* is an ambitious attempt to draw together many of these debates, and to work towards an overarching theory of why and when the law should intervene to regulate sexual relationships between adults. In this book, Schulhofer seeks to establish some unifying principles and to sketch out their possible application in different circumstances. He does so by discussing numerous examples drawn from the case law of various US states. Schulhofer's critique of those cases is the starting point for the formulation of principles he hopes will be of general application. He begins the work with a discussion of rape law, working out his arguments through discussions of decided cases. From those beginnings he moves to analysis of the underlying principles and then back to particulars through application of those principles in other contexts.

The central theme of the book is that we have misconceived the nature of the interests at stake in regulation of sexual behaviour, and have protected only one aspect of these interests at the expense of the other. Schulhofer argues that the current debate lacks a developed theory of sexual autonomy, the elaboration of which would improve analysis and reform of the law in relation to sexual matters. Once we accept the premise of *sexual autonomy* as an important interest worthy of legal protection we can begin to consider its implications. The proposed approach would shape law's response to rape, sexual harassment, and sexual relationships within professional relationships and relationships of trust.

In considering Schulhofer's work I wish to explore the concept of sexual autonomy and to ask whether it should properly be viewed as an interest distinct from other aspects of physical and personal autonomy. Additionally, I wish to consider whether this way of conceptualising the issues can offer a more coherent and principled way of approaching the legal response to claims that such autonomy has been violated. In discussing the book I will compare Schulhofer's approach with that of the Home Office Review, which in several areas has recommended changes similar to proposals advanced in Schulhofer's work.

Sexual autonomy has two aspects, Schulhofer argues: the right to choose to have sex and the right to refuse. Currently the law offers greater protection to the first aspect of sexual autonomy than to the latter, because of the way it approaches consent. The presence or absence of consent divides rape or sexual assault from other sexual contact, but this fails to acknowledge concerns about the factors giving rise to consent. Current law typically makes no distinction between consent which flows from mutual desire and consent induced by coercion, fraud, or exploitation of power imbalances. This approach to consent, argues Schulhofer, favours those stronger parties who are able to exploit power imbalances to secure consent to sex. Their autonomy to pursue their sexual desires is relatively free of interference, but the autonomy of the weaker party to refuse unwanted sex is not protected.

A general question which might be raised about Schulhofer's work is: why does he wish to specify a principle of specifically sexual autonomy, as distinct from physical autonomy generally? Why, for example, should fraudulently obtained consent to sex be treated differently from fraudulently obtained consent to medical procedures? The book contains no discussion of autonomy as a general concept, nor does it look at the question of bodily autonomy outside the sexual context.⁴

4 In another work, however, Schulhofer has described his theory of autonomy as a 'thin theory' disconnected from normative judgements about the goals one might seek through the exercise of that autonomy. S. Schulhofer, 'Taking Sexual Autonomy Seriously: Rape Law and Beyond' (1992) 11 *Law and Philosophy* 35 at 42.

Schulhofer does not say whether he considers sexuality to have special characteristics of an emotional and personal nature that put it in a special category; he simply argues that this aspect of autonomy is insufficiently protected in law.

My own instinct is to say that sexual privacy and autonomy do involve matters which are distinct in kind as well as weight from general matters of bodily autonomy. Consent to sexual contact is not, for example, the same as consent to medical procedures, as the social and personal significance of the two activities is radically different. Ordinary assaults are typically only viewed as serious where some physical harm results, despite the principle that even the slightest contact is criminal if unauthorised. With respect to sexual offences physical harm may be highly significant as evidence that the sex was not consensual, but the extent of that physical harm is not as important in assessment of the seriousness of the offence. Rather, the gravity of the offence lies in the violation of psychic and emotional, as well as physical, integrity.⁵ Sexual integrity and autonomy are not identical with proprietary bodily autonomy, but involve both physical integrity and affective dimensions.⁶ Thus the wrong of rape or other sexual violations is more than just the violation of a proprietary interest in bodily autonomy.⁷ It involves a violation of integrity in what Drucilla Cornell terms the 'imaginary domain',⁸ an infringement of the sort of relational and embodied autonomy explored in the work of Jennifer Nedelsky.⁹ The centrality of sexuality to personhood, and its complex involvement in both physical and affective relations, suggests that there are good reasons for retaining a category of sexual wrongs legally and conceptually distinct from other violations of autonomy.¹⁰

But it will not always be clear where the division lies between sexual and non-sexual matters. Admittedly, in many cases there will be no problem deciding whether a particular invasion of autonomy is of sexual autonomy or of bodily autonomy; but one can imagine cases near the borderline, or cases just either side of the boundary, whose differential treatment might raise concern. Consider the controversial decision of the House of Lords in *Brown*,¹¹ concerning consent to S&M activity. It seems a case concerned with sexual autonomy, but it is not clear whether Schulhofer would see it as raising different issues if the acts in question were performed for non-sexual motives. For instance, if a group of religious flagellants might be prosecuted for assault occasioning actual bodily harm, a general theory of bodily autonomy would suggest approaches but a theory of sexual autonomy would not. Likewise if the participants saw themselves as engaging in performance art, or were attempting to make some sort of political statement. It would be helpful to consider whether the sexualised nature of the activity calls for the application of different principles of intervention.

5 I do not wish to dismiss the reality of serious physical violence and harm in some rape cases, but simply want to emphasise that this is not the only harm involved.

6 See Lacey, n 3 above.

7 As conceived by Drippa, n 3 above.

8 D. Cornell, *The Imaginary Domain* (London: Routledge, 1995).

9 J. Nedelsky, 'Reconceiving Autonomy' (1989) 1 *Yale Journal of Law and Feminism* 7.

10 Gardner and Shute have argued that rape is rightly regarded as one of the most serious crimes not because it is harmful in the usual sense of that word but because it is, in the pure case, 'the sheer use of a person' and therefore the paradigm case of (ab)use of another and therefore a denial of their personhood: J. Gardner and S. Shute, 'The Wrongness of Rape' in J. Horder (ed), *Oxford Essays in Jurisprudence* (Oxford: Oxford University Press, 4th edn, 2000) 205.

11 [1994] AC 212, 97 Cr App R 44.

Force, consent, and the law of rape

One of the persistently problematic aspects of rape law is the boundary between consensual intercourse and rape. English law defines rape as sexual intercourse without consent, a definition which sounds simple but is far from simple in its application. As MacKinnon and others have pointed out, consent is more like a continuum than a clear 'bright line' phenomenon.¹² At one end of the continuum is unwanted sex imposed by the use of violence; at the other is fully consensual and mutually desired sex. In between is sex submitted to out of fear or obtained by exploitation of weakness, sex agreed to out of a desire to please a partner, sex engaged in out of a sense of duty or a wish to avoid a quarrel. Different individuals have different views about where on this continuum to draw the boundary. The Home Office Review has adopted the Law Commission's recommendation that consent in sexual offences be defined as 'free agreement'; the Review has gone on to advocate in addition that the law provide a non-exhaustive list of examples of when free agreement is not present.

As Schulhofer explains, in most US states the criminal law of sexual assault normally requires proof that the accused has used physical force or threatened the victim with physical injury. Intimidation or pressure short of such threats is regarded as forceful persuasion which doesn't vitiate consent. In many states, even a woman's¹³ clearly expressed objections will be irrelevant if the man goes ahead and has intercourse without her consent, as long as no violence was used or threatened. English law has no formal requirement of force. In the case of *Olugboja*, it was held that rape occurred even when a woman offered no resistance due to fear, and 'submitted' to sexual intercourse.¹⁴ The Court of Appeal said juries should be told that consent is to be given its 'ordinary meaning', and, if need be, 'that there is a difference between consent and submission; every consent involves a submission, but it by no means follows that every submission involves consent.' The decision is helpful since it entails the recognition that submission induced by fear is not consent; but it does nothing to clarify what the difference between consent and submission might be. The redefinition of consent proposed by the Home Office Review is intended to make it clear that consent must be more than just submission, and that it cannot be based upon mistake, duress, fraud or intoxication.

There is general agreement that violent forced sex with a stranger is rape, but the majority of rapes reported to police and support services do not fit this pattern; in fact, a large number of rapes are perpetrated by men previously known to the victim, whether casually or intimately. Where the incidents go to trial the defence is often one of consent or mistaken belief in consent. When the acts in question take place in private, and there is little evidence of violence, it may be difficult for the prosecution to persuade a jury beyond reasonable doubt not just the absence of consent but also the defendant's lack of belief in consent . . .¹⁵ In other words, the

12 C. MacKinnon, *Towards a Feminist Theory of the State* (Cambridge, Mass: Harvard University Press, 1989) 171–183. See also C. Smart, *Feminism and the Power of Law* (London: Routledge, 1989) 33–34; R. West, 'Legitimizing the Illegitimate: A Comment on "Beyond Rape"' (1993) 93 Columbia LR 1442.

13 I will refer to the accused as male and the complainant as female because this is the paradigm case of sexual assault with which Schulhofer is concerned, and around which the law of rape has developed. However, English law now includes the rape of a male in the definition of the offence.

14 [1982] QB 320, [1981] 3 All ER 443.

15 For examples, see S. Lece, *Carnal Knowledge: Rape on Trial* (London: Penguin Books, 1997). A recent study confirms that the rape cases most likely to proceed to trial, and to result in convictions, are those in which there is evidence of physical force: J. Harris and S. Grace, *A Question of Evidence? Investigating and Prosecuting Rape in the 1990s* (London: Home Office, 1999).

law tends to operate as though physical resistance by the woman were a prerequisite of a successful rape prosecution. A number of US states have also introduced reforms designed to remove the requirement of force, but in the absence of a clear definition of consent it appears that they too have seen juries unwilling to regard anything less than violence as indicative of the absence of consent.

Schulhofer argues that rape and sexual assault law has failed to evolve as have criminal laws protecting interests in property. He cites by way of example the common law of theft, which in the sixteenth century regarded the removal of goods with force as an essential part of the offence. As criminal lawyers and law students know, the current law of theft no longer requires any force at all, and the offence may even be perpetrated with the apparent consent of the rightful owner. Such 'consent' will not render the taking of property legal if it has been obtained fraudulently or by coercion. It is Schulhofer's argument that a radical rethinking of the rationale for sexual assault laws is needed, and a concomitant modification of current approaches to consent in this area of the criminal law. Just as we now acknowledge that theft may occur where 'consent' is procured by unacceptable means, so must we develop a more sophisticated analysis of the question of consent in the area of sexual behaviour. To make consent *simpliciter* the boundary between criminal and legal sex is approach the question from the wrong angle, because it means ignoring the real normative question of whether the means used to secure consent or submission are acceptable. In this respect the law of rape and sexual assault lags behind the law of theft, which no longer treats the consent/non-consent dyad as determinative of legality.

One of the strengths of Schulhofer's work is that he acknowledges the difficulty and complexity, the 'messiness' of negotiating sexual intimacy. If a woman on a date initially says 'no' but her male companion persists in the belief that she could be persuaded to change her mind, at what point does his persistence become criminal? If she says 'yes' some hours later, is that the end of the matter? A simple focus on the presence or absence of consent might lead to the conclusion that the ultimate consent was determinative. Schulhofer invites us to consider not simply the eventual statement of consent, but also the acceptability of the means used to procure it. If the man charms his companion and engages her interest sufficiently to change her mind, no wrong has been committed. If, however, she complies after unacceptable pressure, the fact that she said 'yes' should not preclude a finding that a wrong has taken place. The next part of Schulhofer's argument, therefore, is a discussion of just what means of procuring consent should be regarded as impermissible.

A workable theory of consent in sexual matters must acknowledge the problems of misunderstandings and mixed messages, of complexity, confusion and ambivalence. Miscommunication about sexual matters is not unusual; rather, it seems depressingly common. Silence about these matters will not make them go away; as Schulhofer observes, behaving as if consent were unproblematic does not render it so, but merely favours certain individuals (the man who ignores the signs of non-consent) over others (the woman who is inarticulate or reluctant to resist with physical force). And the messiness and confusion is not a reason for failing to take sides, but rather a reason for so doing in as clear and principled a way as possible (pp 65–66). Just as the law sets rules for determining property rights in complex and uncertain areas, so it must also set boundaries in the area of sexual rights. Schulhofer favours a positive requirement of clearly expressed consent, thus shifting the risk of misunderstandings from the complainant to the defendant.

Clearly, the requirement of physical force is unduly narrow in an era which recognises the effect of many other sorts of pressure, from economic duress to exploitation of positions of trust. Some feminist scholars have argued that the solution lies in moving from a focus on force to a focus on *power* as the key concept; power which may be other than physical or violent in the narrow sense. Just as the law addresses abuses of power in areas of commercial and property rights, it could also acknowledge the phenomenon of power abused to obtain sex, by treating this as obtaining sex by force. Schulhofer considers but rejects this approach because he thinks it expands the definition of force too far and because it still fails to address directly the need to enhance sexual autonomy. Rather than distort the meaning of 'force', it is preferable to acknowledge that force is not the only unacceptable means of obtaining consent. The next question, therefore, is what these various unacceptable means have in common.

The answer can be found in chapter 6, 'The Missing Entitlement: Sexual Autonomy'. In this chapter Schulhofer sets out his central arguments. He argues that protection of autonomy is linked to protection from force but is conceptually distinct and broader in scope. Sexual autonomy, he argues, has been protected only as a by-product of protection against the use of force, and this is an awkward and undesirable approach. In contrast, the legal recognition of other interests is not so limited. The law protects us from many forms of interference with our autonomy in deciding how to use or dispose of our property. Our autonomy with respect to property interests (or with employment rights, privacy rights, or political rights) is, broadly speaking, protected from unacceptable interference with meaningful choice. That interference may be improper because it involves stealth, force, deception or breach of trust. These are not all treated as 'coercive', but simply seen as various ways in which self-determination may be impermissibly infringed. But sexual autonomy, he argues, is generally protected only from coercive interference with self-determination, much as if it were restricted in scope to a sort of 'sexual robbery' (p 101). Other flaws in consent are usually overlooked.¹⁶

Any autonomy-based approach must begin by accepting that we all experience numerous limitations upon our autonomy in a variety of areas, and that only some of these can be regarded as illegitimate. Background conditions constraining choice must be distinguished from autonomy problems due to individual behaviour, and in Schulhofer's view this distinction is not properly acknowledged in the work of certain prominent feminists such as Andrea Dworkin and Catharine MacKinnon. On the other hand, writers such as Camille Paglia and Katie Roiphe who stress the need for women to take responsibility for their 'choices' go too far when they seem to suggest that we need not concern ourselves with social pressures to engage in unwanted sex (p 110). The wrongfulness of background conditions may be conceptually distinct from the wrongfulness of individual conduct, but that does not mean it should be overlooked entirely. Context is important.

What, then, does sexual autonomy mean? For Schulhofer:

'Fully autonomous choice requires mental capacity, awareness of the available options, adequate information, and freedom from outside interference with the process of choice. When one individual interferes with another person's ability to choose freely in a sexual encounter, legal intervention is appropriate even in the absence of force, just as it is when an individual impairs another's autonomy in relation to property rights, privacy, and the other entitlements that the law already recognizes.' (p 111).

16 The obvious exceptions are in the areas of statutory rape or intercourse with those incapable of giving consent at all (owing to mental deficiencies or intoxication).

Elaborating further, he says sexual autonomy has three dimensions. The first two are 'an internal capacity to make reasonably mature and rational choices, and an external freedom from impermissible measures and constraints.' The third is 'the separateness of the corporeal person' (p 111). As an initial statement this seems commendable, but its application to law is far from self-evident. Schulhofer says these general principles must be developed and applied with due recognition that such general concepts as consent and coercion must be addressed in different contexts, the facts of which must be given due consideration. Protection of autonomy means that consent cannot be regarded as a simple matter dividing legitimate from criminal sex; instead, we must consider the preconditions for valid choice in the particular circumstances of the case. The central question in each case, he contends, is whether there was a meaningful choice which represents an exercise of the sexual autonomy of the individual. In the rest of the book Schulhofer works through the practical application of his approach and tries to show how it may offer a more principled way of tackling these difficult and controversial issues.

One notable omission of the work is any in-depth discussion of the first aspect of autonomy identified by Schulhofer: the internal capacity of the individual to make reasonably mature and individual choices. The matter arises as part of the discussion of certain types of relationships of trust, where one party may be vulnerable, but there is no separate discussion of difficult autonomy issues concerning vulnerable adults generally. Schulhofer's model statute states that consent is not freely given, and therefore not true consent, if the person who gives it is 'physically helpless, mentally defective, or mentally incapacitated' (p 283). The potentially problematic aspects of such an approach are explored in the Home Office Review. The Review devoted considerable energy to discussing the position of vulnerable adults, as did the Law Commission. It was acknowledged that a blanket denial of capacity to consent may be inordinately broad and consequently deny the sexual desires of the disabled. Conversely, a provision intended to give recognition to the positive sexual autonomy of the disabled may deny vulnerable people legal protection from exploitation.¹⁷ The Law Commission's discussion of the issues throws up numerous illustrations of the difficulty of balancing the protection of both positive and negative sexual autonomy. Few would wish to see two mentally disabled adults treated as criminals for choosing to enter into a sexual relationship, yet it is important that the vulnerable not be prey to sexual exploitation. English law currently makes it a criminal offence to have sexual intercourse with the severely mentally disabled,¹⁸ but does not automatically deem them incapable of giving a consent operative in the context of rape law. Thus, in one case described by the Law Commission, a young woman with a mental age of two or three became pregnant as a result of sex with a male member of the staff in the residential institution where she lived. Although she did not understand what had happened to her and was deemed incapable of consenting to an abortion, the trial judge ruled that she had properly consented to the sexual relationship, as 'she simply had to submit to her animal instincts to be deemed to have consented'.¹⁹ The Home Office Review found a dearth of information about the prevalence of sexual abuse of the vulnerable, but noted that it appeared to be extensive, although there were few prosecutions for the specific offences of sex with a 'defective'.²⁰ The alleged failure of law to protect vulnerable people's negative sexual autonomy

17 Law Comm para 4.8.

18 Sexual Offences Act 1956, s 7.

19 Law Comm para 4.66.

20 HOR para 4.2.5; there were 6 prosecutions and 1 conviction for these offences in 1998.

can be contrasted with the Review's observation that several contributors said care staff who were concerned about legal consequences went too far in denying the mentally impaired the right to consensual sexual relationships.²¹

This seems a clear illustration of Schulhofer's thesis that the bright-line approach to consent is not an effective way of protecting sexual autonomy. Using consent as the standard of legality in such contexts is inadequate because it invites courts to overlook difficult issues of autonomy. The Law Commission has suggested a test for capacity to consent in sexual cases which requires the party giving consent to understand the nature, implications and reasonably foreseeable consequences of the act and to be able to make a decision about whether or not to consent. This would protect the negative sexual autonomy of persons like the young woman in the case described, but could create a risk of criminal liability in other situations where it might be thought less appropriate. To cite another example discussed by the Law Commission: a husband and wife might continue to have sexual relations although one of them suffers from a progressive form of dementia such as Alzheimer's Disease. Although the person might lack capacity to consent according to the proposed test, many would find the idea of criminal sanctions unacceptable. The Law Commission's solution is to suggest that certain exemptions could be created for situations where there was apparent consent and the circumstances were non-exploitative. The difficulty of trying to determine just what such circumstances might be is obvious, but the proposal represents a genuine attempt to wrestle with the complex and difficult autonomy issues discussed by Schulhofer. One alternative is suggested by the Home Office Review, which recognised that vulnerable people might be in need of protection even though they were above the threshold for capacity to consent; the Review suggested that the law should prohibit sexual relationships within relationships of trust and caring relationships, as these present the greatest risk of abuse. The Review's discussion contains a real recognition that the internal and external aspects of autonomy are intimately connected; those who are vulnerable due to illness or disability are also likely to be in situations where they are at risk of being subjected to unacceptable external pressures. Unfortunately this book does not explore the difficult questions of autonomy's internal aspects, but focuses almost exclusively on autonomy constraints external to the individual: coercion, fraud, and abuse of power in relationships of trust or authority. The first of these, coercion, is the subject of a thoughtful analysis of what is acceptable and unacceptable when one party indicates that certain consequences may follow the granting or withholding of consent to sexual relations.

Offers and threats – sexual coercion and sexual bargaining

One obvious problem with expanding the concept of coercion beyond threats of violence lies in finding a clear and generally acceptable line between permissible and impermissible pressure of a non-violent sort. Bargaining is acceptable but coercion is not; however, that statement is of little help in determining where one stops and the other begins. Here Schulhofer looks to the economic model, but with reservations. He acknowledges the appeal of an economic or property-based approach, which has attracted Posner and others,²² but claims there are fundamental

21 *ibid.*, para 4.2.3.

22 R. Posner, *Sex and Reason* (Cambridge, Mass: Harvard University Press, 1992); Dripps, n 3 above (arguing that sexual autonomy should be grounded in a theory of property in the self, and violations of this interest categorised as a form of sexual theft).

differences between sexual encounters and most economic transactions. Sexual encounters in violation of autonomy are, he argues, more potentially devastating to the individual than are most violations of property rights. This suggests we might wish to impose liability more readily in sexual matters. But motivations for sexual exchanges are more ambiguous and complex than most economic transactions, and less externally measurable:²³ this gives us reasons to hesitate before imposing liability (p 117). We cannot simply transplant notions of coercion from the economic to the sexual context, nor can we adopt without adaptation similar concepts from laws protecting other personal interests. Rather, the central question must be what kinds of pressures are impermissibly coercive in the context of sexual matters, with parallels from other areas relied upon where useful.

Schulhofer begins with the proposition that threats are inherently coercive and illegitimate, and some offers may be illegitimate either because of their coercive effects or for other policy reasons (p 118). If an employer promises a promotion in exchange for sexual favours, that may resemble an offer but amount to a threat because it suggests that the employee may not be promoted if the demand is not met. The distinction between threats and offers is, in Schulhofer's view, that 'A threat ... is a proposal to make a person worse off than she has a right to be. An offer is a proposal that contemplates making the person better off, in return for her taking the action requested' (p 120). The key to the distinction is the rightful position of the recipient *absent* the proposal. Intellectually attractive when cast in this abstract way, its application becomes less certain in the context of some sexual matters. As Schulhofer is quick to explain, the contentious question becomes that of the correct perspective to adopt when assessing the controlling expectations (p 122). He offers the example of a male student who says he will stop dating his partner unless she agrees to have sex with him. One view is that she has no right to a continued dating relationship, and the ultimatum is therefore not a threat because it leaves her no worse off than she could legitimately expect. Another view is that she may have a legitimate expectation that sexual involvement will be negotiated through mutual discussions, not made the subject of a unilateral demand. On balance, though, the overriding view is that society will not compel people to remain in personal relationships (even marriages) they find unsatisfactory. And so the proposal is not a threat, because it leaves the female student no worse off – the male student was entitled to stop dating her for any other reason, or for no reason at all.

In the employment context, the question of sexual harassment highlights the difficulty of distinguishing between legitimate and unacceptable offers. We can see that a demand for sex is a threat if the employee who refuses will be dismissed without good cause. But if there are grounds for dismissal the employee has no legitimate expectation of keeping her job. If her boss says he will not dismiss her if she complies with his request for sex, the intuitive response is that such a proposal is still unacceptable. Schulhofer tries to determine just why this should be so. Coercive inducements, as Schulhofer sees these, are illegitimate interferences with sexual autonomy. He views them as analogous to demands for financial kickbacks; the party paying such a bribe is a victim of extortion even if advantages are

23 Schulhofer gives an example of a large man and small woman meeting on a park bench and conversing briefly; if the man asks the woman to hand over her money and she complies, most people would assume she was motivated by fear. If, however, he asks for sex and she complies, it appears many jurors would find at least a reasonable doubt that she consented for other reasons. This may be due to the intuition that in the monetary exchange it is difficult to see any benefit to the party handing over the cash, while the possibility exists that both parties could derive pleasure from the sexual encounter.

obtained in exchange. In both cases the individual presented with the offer is denied the opportunity to compete on fair terms, 'to seek a decent income and professional fulfillment by pursuing the normal activities appropriate to the setting' (p 149). In the case of the kickback the price is financial, but in the case of sexual offers the individual is asked to relinquish sexual privacy and bodily integrity. What Schulhofer does not address explicitly (although it is implied in his discussion) is the idea that such illegitimate offers damage not only their recipients but also those who might otherwise have secured the advancement or opportunity. Focusing solely on the harm to recipients of such offers poses problems when the impugned transaction is initiated by the party offering the bribe or sexual favours. As Schulhofer acknowledges, where custom and expectations are that such exchanges will lead to preferential treatment there may be no explicit offer by the party with the power. In Schulhofer's view this situation can only be regarded as coercive when the person in power provides illegitimate benefits and knows the victim is afraid of suffering harm if sex is not offered. Otherwise, the arrangement is illegitimate but not a coercive interference with autonomy. The distinction seems too fine, though, when one tries to imagine the situation of the less powerful party who might feel equally exploited whomever was the party initiating the proposal. After all, the most feared tyrants need not state their desires explicitly, as others rush to please them. Problems arise because what *feels* coercive to the less powerful person may seem very different to the more powerful party. It may be appropriate, however, to regard the two situations as distinct in terms of criminal liability, even if both seem to some degree coercive.

Even if a proposal is a threat, of course, the law generally views as coercive (and vitiating consent) only those threats which a person of reasonable firmness could not be expected to resist. This is true in the criminal law of duress and in disputes about contractual rights. Schulhofer says those restrictions serve particular needs (in the commercial context, the need to permit renegotiation of agreements; in the criminal context, the need to limit the circumstances in which defendants can escape liability for their harmful acts) not present in the context of sexual matters. He sees the better analogy as that with extortion – the law does not absolve the perpetrator of responsibility simply because the victim had alternatives to submission, such as going to the police or engaging in self-help. The wrongful threat amounts to extortion even if the victim had other choices. Schulhofer takes a strong position here (p 132): 'Conduct that forces a person to choose between her sexual autonomy and any of her other legally protected entitlements – rights to property, to privacy, and to reputation – is by definition improper; it deserves to be treated as a serious criminal offence.' Such an approach could, if applied rigorously, significantly expand the scope of criminal law in this area. It could, for example, render criminal things currently dealt with through employment law as sexual harassment.²⁴ In the English context, it might mean either the creation of a completely new offence, a radical expansion of the definition of rape or a new approach to existing extortion offences.

Students of English criminal law will recall the case of *Clarence*,²⁵ in which a wife claimed that her consent to sex with her husband was invalid because he had lied to

24 Schulhofer does state that the criminal law need not be the sole or main vehicle for protection of sexual autonomy, but much of the discussion throughout the book does focus on criminal sanctions. This is unfortunate, as there are likely to be many readers with some sympathy for the general approach of the work who will nonetheless be wary of expanding the scope of the criminal law in this area.

25 (1888) 22 QBD 23.

her by pretending he did not have a venereal disease. In that case the general rule was affirmed that only a misrepresentation as to the 'nature and quality of the act' would vitiate consent. The court expressed concern about the consequences should any less restrictive approach be adopted – imagining the men who would become rapists simply because they had lied about their marital status, their honourable intentions, or their feelings for the woman in question. Schulhofer is similarly cautious, noting that in sexual relationships complete honesty is neither expected nor universally desired. His instinct is that most deceptions inducing consent to sex should be left unregulated by law, with possible exceptions for impersonation or misrepresentation as to significant health risks like HIV (p 159). The seriousness of HIV appears to put it into a category different from other types of deception, as the Supreme Court of Canada recently held when ruling that consent was not an effective defence when an HIV-positive man knowingly had unsafe sex with several women.²⁶ But surely with respect to HIV the problem is not essentially one relating to sexual autonomy, but is about transmission of disease, whether that transmission is through sexual contact or by sharing of hypodermic needles. Current English law holds that even wholehearted and informed consent is ineffective as a defence to the infliction of serious harm, although whether transmission of HIV falls within this category is debatable. The approach of the proposed new Offences Against the Person Bill would be to treat intentional transmission as giving rise to a prosecution for a general and non-sexual offence.²⁷ This seems right, although there might be a sexual offence component to the wrongdoing in addition.

But it is not only fraud sufficient to vitiate consent which can give rise to criminal liability in English law. It is an offence to procure a woman by false pretences for sexual intercourse anywhere in the world.²⁸ The Review has recommended reshaping the offence to one of 'obtaining sexual penetration by threats or deception in any part of the world' (Recommendation 14). It would apply to situations where consent was not *absent* (the Review thought the categories of fraud vitiating consent should remain narrow), but was improperly obtained. The Law Commission also found merit in the idea of a lesser offence of obtaining consent by deception, and in retaining narrow limits on fraud vitiating consent. Neither body, however, addresses the concerns Schulhofer raises about our expectations of honesty in sexual matters and the difficulty of drawing lines between acceptable and unacceptable deceptions. The Review suggests that the type of deception which might lead to liability would include reckless as well as deliberate deception, and deception by words or conduct as to any fact or law, including present intentions.²⁹ This could render criminally deceptive a wide range of misrepresentations from the obviously reprehensible to the all-too-commonplace. Its breadth is worrying, and a legislator considering such an enactment would do well to read Schulhofer's analysis of the matter. As he observes, in intimate matters distinctions between 'welcome storytelling and unwelcome lies' (p 158) may be especially elusive. We might not all wish to treat as unlawful the deception of a spy who feigns love in order to marry – in 1999 a Florida woman initiated a civil action against the Cuban state, alleging that her husband was a Cuban spy who married her to gain cover rather than for love, and that therefore his fraud vitiated her consent to sex.³⁰ It is even less clear that we would want to criminalise

26 *R v Cuerrier* [1998] 2 SCR 371.

27 Available through the Home Office website at: <http://www.homeoffice.gov.uk/cpd/sou/oapdb/htm>

28 Sexual Offences Act 1956, s 3.

29 The Review considered the definition of deception for the purposes of the Theft Act.

30 S. Krum, 'Woman duped by Cuban agent sues for rape' *The Globe and Mail* (Toronto), 24 August 1999.

the bad behaviour of the partner who lies by denying his engagement to marry another. To take an absurd example, consider the character in the film, *Educating Rita*, who lies to her husband by concealing her use of contraception. If he said he was induced to have sex by means of this deception, few would think a criminal sanction appropriate.

Supervisors, teachers, doctors and lawyers: power and consent

Chapters nine through eleven of the book contain discussions of how sexual autonomy might be protected in the context of relationships where a duty of trust or power inequalities raise questions about the legitimacy of consent to sex. In these contexts it is not unusual to find that individuals are mutually attracted; sexual exploitation is not unusual either. As Schulhofer observes, a balance must be struck between the need to protect vulnerable individuals from exploitation and the need to protect the individual's autonomy to engage in voluntary and freely chosen relationships. As employers become increasingly concerned about the risks of litigation arising from sexual harassment, it is unsurprising that some have reacted by attempting to impose no-dating policies only slightly less strict than military anti-fraternisation policies. But these deprive employees of the freedom to enter into fulfilling and non-coercive relationships with co-workers. It is difficult to devise a policy which will target only exploitative or abusive workplace relationships while permitting acceptable ones to flourish. Schulhofer considers the alternative model of banning sex between supervisors and subordinates, but rejects it as unduly restrictive, concluding that perhaps the best approach is something close to the current definition of sexual harassment as unwanted advances, but with a greater willingness on the part of courts and tribunals to accept that propositions are unwanted (either crude initial proposals or less vulgar approaches which follow an earlier rejection) (p 187).

In the end, then, his autonomy-based approach does little to help arrive at a new way of addressing this problem. Employment law has already acknowledged that some sexual activity in the workplace can constitute a wrong despite 'consent'; the thorny question is just where to strike the balance between protection and paternalism.

The teacher-student relationship is also one characterised by a power imbalance. US law currently treats sexual advances by teachers or academics in much the same way as sexual harassment in the workplace;³¹ there must be either a 'quid pro quo' proposition which the student found objectionable, or a 'hostile environment' characterised by unwelcome advances, communication by the student of this unwelcomeness, persistent or severe harassing conduct, and negligence by administrators in failing to prevent the conduct. Of course, almost all schools prohibit teacher-pupil sex, and many universities have codes of conduct regulating or prohibiting sexual contact between staff and students. The approaches taken to such relationships vary widely, with some institutions completely prohibiting them and many others declining to adopt any policy at all. Schulhofer argues that a university has a genuine interest in avoiding the conflicts of interest presented by faculty-student liaisons, but again concludes that a sweeping general prohibition would unjustly limit sexual autonomy. A more narrowly tailored prohibition

31 The discussion does not address sex where the student is below the age of consent, or where the teacher uses physical force. In such cases the incident is already criminal.

however might acceptably ban all sexual contact between an academic and a current student.

England may soon see greater legal intervention in this area. The Sexual Offences (Amendment) Bill would render unlawful any sexual activity between certain persons in positions of trust and any person under 18 subject to their authority, regardless of consent. The reasoning behind the legislation can also be found in Home Office guidelines for preventing abuse of trust by those caring for young people³² and vulnerable adults.³³ The guidelines state that there are two reasons why sexual relationships are inappropriate between parties in a relationship of trust. One is that the 'professional' relationship of trust would be altered; the more important one, however (at least, the one given most prominence in the guidelines) is that 'A sexual relationship itself will be intrinsically unequal within a relationship of trust and is therefore unacceptable' (p 5). But almost every sexual relationship (in fact, almost every human relationship) involves some degree of inequality, as Schulhofer observes in his analysis of sex in the workplace. That in itself need not rule out the possibility of a relationship; the central question, in his analysis, is whether that inequality renders autonomous consent impossible. There can be no simple universal answer to this question: the extreme view, that equality is a precondition to acceptable relations, conflicts with the experience of those who have found valuable and desired relationships within an imperfect and unequal reality. Equally, to dismiss the problematic nature of power differentials between sexual partners is to turn a blind eye to serious issues of exploitation and abuse. The relationships addressed in the Bill and Home Office guidelines are ones of such inequality that an outright ban on sexual relations must be justified despite the risk that it might rule out a few genuinely non-exploitative relationships. But there are many relationships of inequality where the need for protection is not quite as clear, such as those between professionals and their clients or patients.

Professionals in positions of trust may be subject to professional discipline when they have affairs with clients, but civil remedies are problematic and criminal sanctions virtually non-existent. This is the case despite the fiduciary nature of the relationship. However, in a few US states laws have been introduced to grant civil remedies for sexual harassment within the context of a range of relationships, including not only physician-patient and lawyer-client but also landlord-tenant and trustee-beneficiary, among others. The most problematic encounters occur in the context of patient-therapist relationships, as the patient's consent will almost always be flawed, there is a clear conflict of interest between the roles of therapist and lover, and the potential for harm and abuse is great. The small number of patients who might be capable of entering into a beneficial and freely chosen affair is so small that their interests are outweighed by the need to protect the rest; in this context Schulhofer advocates a complete ban on sexual contact. In fact, he sees the potential for abuse as so serious that criminal sanctions may be justified in addition to professional discipline and civil remedies.

Doctors, dentists and lawyers are also in positions of trust. Problems of power abused, violation of trust, or conflict of interest may all give rise to concern, but are not necessarily present in all relationships of this sort. The circumstances are crucial: in the medical context, factors such as whether the treatment is short-term or ongoing, whether it requires emotional intimacy, whether the encounter occurs

³² Meaning over the age of consent but under 18 years of age.

³³ *Caring for Young People and the Vulnerable? Guidance for Preventing Abuse of Trust* (London: Home Office, 1999).

in the doctor's office or elsewhere (p 235). The American Medical Association prohibits all sexual contact between physician and patient, even if they are husband and wife or were sexual partners before the medical relationship arose. Schulhofer argues this is too broad. His qualms seem odd, though, if one considers that in most such cases the spouse-patient will be free to seek the services of another physician. As he says (p 238), 'Legal intervention is normally unnecessary when a person subjected to unwelcome advances can simply reject them and walk away.'³⁴ In conclusion (pp 238–239), Schulhofer argues that legal remedies should be expanded and professional restrictions narrowed with respect to sexual relations between doctor and patient. Sexual contacts in the office should be subject to an unqualified ban, as should sexual initiatives that exploit information or emotions expressed in trust, and sexual advances where the patient's ability to refuse is inhibited by the medical relationship. But consensual relationships not tainted by a 'significant potential for abuse' should be permitted, and Schulhofer thinks these concerns 'can be captured in workable rules that are not hopelessly abstract or indefinitely elastic' (pp 238–239). What the rules might look like in practice, however, is not explained. The Home Office Review has also advocated banning all sexual relationships between doctor and patient or therapist and client. The Review expressed uncertainty, however, as to whether the ban should apply to physical therapists as well as psychotherapists. They did not examine the next category of relationship considered by Schulhofer: that of lawyer and client.

The US ethical rules applicable to lawyer-client relations vary dramatically from state to state. Some state bar associations ban all lawyer-client sex, while others refuse to proscribe it under any circumstances.³⁵ Here, Schulhofer favours a bright-line prohibition of all lawyer-client sexual contact in the areas of practice he feels pose the greatest risks of coercion and exploitation.³⁶ For other areas of practice, he advocates rules that declare sexual relations to be unethical whenever they involve quid-pro-quo demands, sex in the office, use of confidential information, or any other form of coercion or undue influence. Of course, his earlier discussion has asserted that such factors make consent flawed even outside the context of the professional relationship. In fact, the problems discussed in the lawyer-client context might well arise in numerous other professional and personal relationships: inequalities of power are the norm, not the rule, and they bring the risk of exploitation. Schulhofer's arguments about professional ethics could as well be applied to a wide range of relationships.

Concluding remarks

Schulhofer offers a conception of sexual autonomy which entails aspects of both negative and positive liberty, but the book is dominated by discussion of the

34 Schulhofer is writing in the US context, however, where the restrictions imposed by health care providers and insurance companies may make it more difficult for a patient simply to leave.

35 Of course, the absence of an explicit rule with respect to sexual relationships does not mean that such relationships are unregulated by professional ethics. Rules about conflicts of interest and abuse of trust will apply to these relationships as they apply to other lawyer-client relationships. But Schulhofer notes (p 252) that these general rules are not enough because there is intense disagreement within the profession over whether sexual relationships do involve conflicts of interest, misuse of confidences or prejudice to the administration of justice.

36 These are: divorce, child custody, criminal defence, personal injury claims by plaintiffs and matters of probate, trust and estates. These areas, Schulhofer states, are characterised by 'emotional content, personal disclosure, and difficulty of exit' (p 253).

freedom from unwanted sex, rather than the freedom to engage in wanted sex. This is not surprising, given the relative scarcity of legal limits upon wanted sex between consenting adults. In this book, the positive aspect of sexual autonomy is addressed when it operates to limit measures designed to protect against unwanted sex, as in the ban upon sexual relationships between military personnel. However, Schulhofer seems to view this issue as principally about negative liberty in the sense of freedom from legal restriction of autonomy. The discussion does not proceed further to explore the possibility of a positive form of sexual rights – the question of whether an individual might legitimately claim that he or she be offered sufficient opportunity to enjoy a flourishing sexual life. It is difficult to imagine such a positive liberty in operation, however, except perhaps in the case of disabled people who might seek state-supported access to sexual services not otherwise available. Equally, it cannot be realistic to propose that positive sexual autonomy should be secured by giving individuals the right to demand sex from others, as that would infringe the autonomy of others.³⁷

The idea of autonomy as the central basis of protection is an attractive one. As with many attractive concepts, however, near-universal agreement as to its desirability is not accompanied by similar agreement as to its definition or the best means of securing its protection. Schulhofer takes as axiomatic the desirability of autonomy as a core value, and directs his argument to the proposition that the law has failed to afford adequate consideration to sexual autonomy, which he says is a sufficiently weighty interests as to merit legal recognition. He does not, however, explain with precision just why sexual autonomy should be viewed as conceptually and legal different from physical autonomy generally. Schulhofer compares the protection of sexual integrity and autonomy now with the legal protection of property rights in previous centuries, but does not make the more obvious comparison, between protection of sexual autonomy and physical autonomy/integrity, save in limited sections of the argument. Nor does he explore the affective and social aspects of sexual autonomy, its place in the formation of intimate personal attachments and relationships. He appears content to leave others to argue for the value of sexual autonomy, while he concerns himself principally with the practicalities and principles relating to its legal protection.

Despite its limitations, however, this book makes a valuable contribution to the debate about legal responses to sexual wrongs. It prompts us to think about these issues broadly rather than in isolation, and in doing so compels an inquiry into the nature of the interests at stake. In its attention to decided cases, specific relationships and existing practices it offers a useful working through of the practical consequences of taking sexual autonomy seriously. Thus it offers a supplement to and development of other work discussing sexual autonomy. One may disagree with some of Schulhofer's conclusions about the scope and form of legal intervention but still find many valuable ideas in his analysis.

37 This is the reason Dripps gives for regarding sexual autonomy as purely concerned with negative liberty in the sense of freedom to refuse sex. He does not address the desirability of protecting the positive autonomy interest that Schulhofer values – that is, the freedom to seek out mutually desired and satisfying relationships.

REVIEWS

David J. Ibbetson, *A Historical Introduction to the Law of Obligations*, Oxford: Oxford University Press, 1999, xlv + 307pp, hb £35.00.

Our millennial common law is apt to come to an ignominious end now that the legislative and judicial powers to make and unmake it have been delegated to strangers from a quite different legal tradition. It is therefore some comfort that its fascinating past has attracted the stylish intelligence of so many remarkable scholars: Maitland may soon be in Westminster Abbey, but happily still with us are Toby Milsom, John Baker and Brian Simpson, to name but three. These are now joined by David Ibbetson who has written a common law counterpart to Zimmermann and an obligational companion for Brian Simpson's work on land law.

English legal historians have been in some respects fortunate. Whereas in France interest in the history of private law was defused by the caesura of 1804 – the marvellous muniments of *Les Olim* are only now being studied – and in Germany the schism between Romanists and Germanists made global treatment difficult, English law has been quite continuous and its different components, such as canon law and equity, can be accommodated without too much difficulty, as David Ibbetson shows.

In other respects, however, the task of the English legal historian is more taxing than that of continental writers, for his material comes preponderantly from the bench rather than the study. This was bound to be so, given England's perennial (and continuing) distaste for disinterested thought: whereas the continent had 59 universities by 1500 and little Scotland already had four by 1603 when its king came South to rule the country which had decapitated his mother and was about to decollate his son, it took England more than 200 years more to produce its third. The English legal historian's material is thus procedural rather than substantive, argumentative rather than propositional, practical rather than conceptual, *dicta* rather than *ratio*, resting on allegations not findings. Often we have only the pleadings, not even the facts, and very often not the outcome, for what the jury did and why is generally unknown and unknowable. As the author says '... the principal losers are the legal historians who seek for clear answers to questions that were never explicitly asked' (p 57). Imagine trying to cope with modern case-law if the actual disposition of the case were secret, and all one had were the arguments of counsel and the interventions of the judges.

How specifically English was the law and its development? Certainly the author, starting out with a brief description of the relevant Roman law, observes that 'The friction between [native ideas and sophisticated Roman learning] was a prime force for legal change, and has remained so right up to the present day', but this is hardly borne out by the text. Lord Holt CJ admittedly invoked Roman law as regards bailments (a useful common law concept unknown to Roman law) and Lord Blackburn was forever flashing his learning, helpful or not, but it is not every reference to foreign law that is significant. For example, the author says of one case that 'the analogy with the Scots law based on the *condictio causa data causa non secuta* was drawn' (p 275); all that was said, however, was 'An action at law lies for money had and received to his use, which is expressed by the Scotch law, *causa data sed non secuta*'. This is allusion, not analogy. Again, one would like some warrant for the following: '... by 1700 it was established as a rule, almost certainly introduced by analogy with Roman law, that where the plaintiff had been directly injured by the defendant's act the correct action was trespass, and

where the injury was merely consequential case was appropriate.' Similarities may arise spontaneously, but analogies have to be drawn by someone. Who did the drawing? After all, when legal solutions coincide, it may be *just* a coincidence.

One must conclude that there is precious little evidence of the transplants of which comparatists never cease talking these days. One important device which might have been invoked as a transplant (though it is simply a similar solution to a cognate problem) is the conditional bond. In Rome it was standard to stipulate that one's debtor pay a penal sum if he failed to perform the service which was the real object of the transaction, and it was also used to sideline the doctrine of privity (another common feature!) by having the debtor promise to pay if he failed to render a prestation to a third party. Rather surprisingly, Ibbetson writes: 'Unlike in Roman law, there was no difficulty in allowing third parties to be affected indirectly. This was most easily achieved by the conditional bond ...' (p 77).

Our author is very much on his guard against undue dogmatism, speaking, for example, of 'the rule (if rule it was)' (p 41). In his masterly review of the substantive law of tort in the Middle Ages ('inherently messy, though perhaps no more messy than its twentieth-century counterpart' (p 57)), he is right to say both that 'it is hard to avoid the conclusion that liability in the writ of trespass was *prima facie* strict' (p 58) and that 'Liability in tort in the Middle Ages ... was suffused by some idea of fault.' The author is also ever conscious of the fact that conceptual clarity and practical justice may not concur: thus 'From a purely intellectual point of view, such lack of concern with the taxonomy of legal claims and the identification of the whereabouts of ownership smacks of a high degree of sloppiness of thought; from a practical point of view, it is beyond reproach' (p 110-111), and 'In so far as this achieved a measure of practical justice, it did so only at the cost of a breakdown in conceptual coherence' (p 247). Even so, it is possible, as the author shows, to indicate with clarity what the confusion was:

The potential for confusion was enormous. *Indebitatus assumpsit* was in substance an action to recover a debt but in form an action for damages: debt on a conditional bond was the reverse, in form an action to recover a debt but in substance an action for damages; and special *assumpsit* was both in form and substance an action to obtain damages. ... Even at the moment when the Common law had the potential to recognise a truly general contractual remedy, it was compromised by procedural twisting (p 151).

The history is brought right up to date, with interesting discussions, for example, of the development of unjust enrichment and the question of the liability of public bodies; the passage on the disappearance of the civil jury (p 188-189) might have been reinforced by a reference to *Ward v James* [1966] 1 QB 273. Compendious though it is, the book is not long, and this is all the more remarkable in that the style is not at all dense, but easy and flowing. The text is replete with well-chosen examples, and the footnotes are informative and stimulating: for example, 'A sufficiently high proportion of defendants who waged their law failed subsequently to make it that we may be sure that it was not just a cynical formality' (p 32).

This indicates how carefully the author has trawled the original sources, and he comes up with some fascinating catches, such as the fifteenth-century case in Chancery where the plaintiffs' scheme for charging the townspeople for a sight of their ostrich was frustrated when the defendants got hold of it and paraded it through the town for all to see for free (p 68).

As pleasurable as it is informative, as balanced as it is intelligent, this volume is an invaluable addition to a distinguished literature.

Tony Weir*

* Fellow of Trinity College, Cambridge.

Arthur Ripstein, *Equality, Responsibility and the Law*, Cambridge: Cambridge University Press, 1999, xii + 295pp, hb £37.50.

This book takes one to the heart of recent legal-philosophical debate about liability and responsibility in tort and criminal law. The rules of liability constitutive of these areas of law and the conceptions of responsibility they invoke raise obvious questions of fairness and justice: can it be right, for example, to hold a defendant to a duty of care they could not actually have achieved? Is it fair that tort law protects those with 'egg-shell' skulls while ignoring those with unusual sensitivities? Should those who genuinely but unreasonably believe consent is present escape conviction for rape? It is questions such as these that Ripstein seeks to answer. While his actual answers to these and other similar questions are not particularly *outré* or challenging, their basis is quite distinctive. Some jurists have sought to explain and justify the liability rules of tort and criminal law on either economic or utilitarian grounds, while others have tried to show that these rules derive from some or other non-utilitarian moral conception of responsibility and its cognates – fault, blame and praise.

There are severe problems for both approaches. Utilitarian-cum-economic approaches struggle to accommodate many important features of the law, such as tort law's bipolar structure. If the concern is with utility or efficiency, why restrict our focus to what went on, in the past, between these two particular parties? For example, it might at the moment of adjudication be more efficient for someone other than the defendant to compensate the plaintiff's loss but, if so, existing systems of tort law are a poor way of discovering this. Approaches that derive liability rules from non-utilitarian conceptions of responsibility face a similar difficulty, in that there is often either legal liability in the absence of moral responsibility or no legal liability in the face of moral responsibility. Accounts of moral responsibility that rest legal liability upon a combination of what defendants actually foresee or intend and what is under their control, face the difficulty that legal liability often arises when defendants did not foresee or could not completely control the harm to the plaintiff. By contrast, those accounts of moral responsibility that focus only upon causation, recommending it as the sole basis of legal liability, have difficulties explaining far from rare situations in which defendants undoubtedly cause harm but are not legally liable. Ripstein rejects these two accounts of moral responsibility and their injunctions for legal liability. The latter approach, associated with the work of Richard Epstein, is shown by Ripstein to be fatally flawed by virtue of its inadequate account of causation (pp 32–46). The former approach, often articulated in discussions of moral luck and elucidated most trenchantly by Thomas Nagel, seems unacceptable to Ripstein only because it conflicts with tort doctrine (pp 72–74).

But if these two varieties of moral responsibility are, in conjunction with the utilitarian-cum-economic approach, struck off the agenda, how can normative sense be made of the liability rules in tort and criminal law? Ripstein's recommendation

begins with the idea of a fair division of risks, and explains tort liability in terms of the imposition of additional risks. Those who impose additional risks can be thought of as owning the risks. If those risks ripen into injuries, they own the injuries, and the payment of damages serves to return those injuries to their rightful owners (pp 46–47).

Criminal law is understood in much the same way by Ripstein, save that it 'serves primarily to protect and vindicate fair terms of interaction. Tort liability is appropriate when someone *takes* a risk with the security of others; criminal liability is reserved for the narrower class of cases in which someone *chooses* a risk (or result)' (p 134). However, it is not only the ideas of risk distribution and fair terms of

social interaction that make Ripstein's account distinctive; also significant is the idea of the reasonable person which he uses as the touchstone for assessments of reasonable and unreasonable risk distribution. On the supposition that all have interests in security and liberty, the problem that tort and criminal law raise is the proper way of balancing these interests. My interest in liberty comes into conflict with your interest in security as soon as my actions impact adversely upon you (I pollute your land or steal your car). Whose interest prevails? Ripstein's approach is to 'balance[. . .] liberty and security by constructing an ideal of a representative person [. . . This] approach expresses an ideal of equality, for it aims to protect people equally from each other, by supposing all to have the same interests in both liberty and security' (pp 6–7). The reasonable person is Ripstein's bridge between equality and responsibility, which is crucial since '[o]ne of the central claims of . . . [the] book is that, far from being opposed ideals, individual responsibility and social equality need to be understood together' (p 1).

Sketching the raw materials of Ripstein's argument cannot do justice to the detail and thoroughness of his treatment of tort and criminal law doctrine. He seeks to demonstrate the various ways in which doctrine incorporates and responds to the related notions of fair terms of interaction and the unreasonable imposition of risk. This close engagement with doctrine sets the book apart from much philosophical work in this area and illustrates the value of work that is both legally and philosophically engaged. That is, work rooted in, on one hand, an awareness of the philosophical complexity and difficulty of notions such as responsibility, liability, causation and fault and, on the other, an appreciation of the detail and subtlety of doctrine. Much work in this area ignores one of these dimensions and privileges the other. Ripstein rarely does. The main body of the work – chapters 3 to 7 – concentrates on both, examining the ways in which doctrine illuminates and is illuminated by the ideas of risk distribution and fair terms of interaction. However, an outline of the raw materials of the argument does have the advantage of illuminating its principal moves and, hence, its principal pressure points. The parts of Ripstein's argument that have to bear most weight and which are therefore at most risk of severe stress are: (i) the idea of fair terms of social interaction; (ii) the idea of a reasonable, or fair, distribution of risks; and (iii) the mechanism of the reasonable person.

The first two notions are advantageously deployed and illuminate much. However, obscurity threatens because both are often run together with a conception of equality that, at best, is vague and, at worst, undermines the ideas of fair or reasonable distributions of rights, resources and risks. The preliminary point here is that what makes a distribution fair and reasonable (for present purposes let us regard them as synonyms) need not make it an equal distribution, in the sense that each has the same share. Hence a fair distribution of risks need not be an equal distribution of risks. But in claiming that '[t]he fault system serves to divide risks at two levels' (p 50), through the duty of care and the standard of care, and that each 'serves to define the equality of the parties' (*ibid*) or 'expresses a conception of the parties as equals' (p 51), Ripstein risks losing the point. For, if the actual division of risks between the parties need not be equal in order to be fair, in what sense does such a division recognise or define the equality of the parties? Hopefully only in the thin sense that each are recognised as being alike or equal because both are risk bearers and imposers. Were a more demanding conception of equality in play, which required that each had equal bundles of risk to bear and impose, then Ripstein would face the difficulty of reconciling this with his acceptance of the basic tenets of historical entitlement accounts of justice (pp 46–44). Such accounts hold that fair moves from an initially fair distribution of rights, risks and resources are acceptable, even though

the result is not a strictly equal distribution. How, then, could tort law express a conception of strict equality? The waters are muddled a little more when we learn that 'the fundamental principle of equality . . . requires that one party may not unilaterally set the terms of interaction' (p 49). However fundamental this principle might be, it is no help in deciding whether parties are equal in the thin sense (both bear and impose risks) or the substantial sense (both bear and impose equal bundles of risk).

Let us be puzzled as to what this has to do with adjudication in tort law, Ripstein has an answer: '[a] fair distribution of risk is a general matter, but courts are called on to decide liability in specific cases. The reasonable person provides the standpoint from which the general distribution of risks can be applied to particular circumstances' (p 56). Hence a great deal of weight is borne by the concept of the reasonable person. Again, the notion of equality is in play and, again, the problem already noted arises: in order to determine whether or not a particular imposition of risk is reasonable, '[w]e look to the liberty and security interests of representative persons – the reasonable person – and protect all equally with respect to those interests' (*ibid*). What does 'equally' mean here? In addition, an objection to the mechanism of the reasonable person highlights its distance from what may be called subjective moral fault. This notion would make assessments of moral and legal liability turn upon the cognitive capacities of the wrong-doer, her capacity to control the outcome in question, and her general circumstances and context. Sensitivity to all these factors surely allows a much finer and discriminating assessment of the wrong-doers' fault than could be achieved by recourse to the fiction of the reasonable person. That notion is a fiction, of course, because it ignores much of what is significant about this specific wrong-doer, her deed and her situation. While Ripstein defends his account against Pashukanis's attack upon the legal form and the legal subject by doubting the coherence of Pashukanis's alternative (pp 254–263), he offers no reply to this related objection. The objection is that the legal subject, understood through the mechanism of the reasonable person, is a morally unacceptable means of assessing fault because it ignores the particularities of the wrong-doer and the detail of her situation. Its power, if it has any, need not depend on the plausibility of Pashukanis's alternative.

Whatever the weight of these quibbles, they cannot detract from the significance and utility of this book. The prose is occasionally dense and oblique – a result, in most instances, of compression – but the argument is unfailingly stimulating. Only the most intellectually incurious could read it without profit.

William Lucy*

Ruth Rubio-Martin, *Immigration as a Democratic Challenge: Citizenship and Inclusion in Germany and the United States*, Cambridge: Cambridge University Press, 2000, viii + 270 pp, pb £13.95.

The increase in human mobility in the post-war period has raised many concerns in Western states such as the role of economic migration, the change in national cultures and the issue of race. As the economies of such countries expanded during this period, it became necessary to encourage migration in order to satisfy the demands of

*Department of Law, Keele University.

the labour market. This book is intended as a contribution to the debate over the place of such immigrants who have settled themselves overwhelmingly in Western states with commitments to maintaining liberal and democratic constitutional orders. In particular, the purpose of this book is to defend the idea that there is a compelling need to overcome the democratic legitimacy gap posed by the differentiation of civil and political society that has arisen owing to the growth in international labour migration. Having left their countries of origin immigrants have contributed to the national economies of the receiving countries but have, on the whole, been unable to participate directly in the political system. According to the author, this has resulted in an increase in politically vulnerable and disenfranchised communities of individuals who contribute much to the social and economic life of their countries in which they live. That access to national citizenship is made difficult in some countries – the author singles out Germany here – contributes to a gap in democratic legitimacy by allowing for the long-term coexistence of people with less than equal political rights. In other words, the post-war processes of economic globalization which have stimulated the mobility of labour across frontiers need to be accompanied by the recognition that democratic legitimacy in the receiving states can only be preserved if the same rights are accorded to those permanently resident immigrants as are enjoyed by the national citizenry.

This book aims to demonstrate why the democratic gap is unacceptable and why it is increasingly important to link the commitment of Western states to liberal democratic principles with the degree of inclusiveness of their citizens. The author seeks to achieve this by advancing the argument that the moral principles underpinning the notion of liberal democracy require full rights of citizenship to be accorded to those immigrants who have been permanently resident for a sufficient length of time. This argument is augmented by an examination of the constitutional jurisprudence of two countries with distinct traditions of immigration, the United States and Germany, that have been compelled to examine the extent to which their constitutions confer upon migrants the same rights enjoyed by national citizens.

The central argument advanced by the author is that those immigrants who are permanently resident in a liberal democratic state should be considered as members of that democracy and therefore share in the sphere of civil equality with the equal recognition of rights and duties. The enjoyment of such rights and duties within the political community of the state remains attached to the recognition of the formal status of national citizenship. Therefore those immigrants who are permanently resident should be recognised as citizens of the state. It is suggested that the period of ten years permanent residency is sufficient to demonstrate to the state that the migrants should be recognised unambiguously as full members of society and not merely members of a temporary labour force. It is contended that this citizenship recognition should cover both legal and illegal immigrants and that it should be automatic and unconditional. Furthermore, immigrants should be recognised as national citizens regardless of whether or not they already enjoy the status of national citizens in another state. In support of this argument the author draws extensively upon the idea of liberal democracy itself. If a state is to fulfil its claim to be democratic then it must exhibit inclusiveness and set limits upon the range of exclusions it may legitimately make. As resident migrants are subject to the laws and political authority of the state, they must be afforded an opportunity to have an equal input into the decision-making process rather than be relegated indefinitely to the limited function of being mere employees. The notion that national citizens alone constitute the sole reference point for the legitimation of the state not only overlooks the short history of nation-states as political structures but also emphasises the ethnic

and cultural elements of citizenship over the political configuration. Rubio-Marín therefore urges that this notion ought now to be abandoned. Furthermore, according automatic and unconditional national citizenship to permanently resident immigrants would have freedom-enhancing implications for both them and the state. The image of such individuals as serving a solely economic function would be replaced by one in which they are encouraged to understand their residency in a more encompassing way. They would be able to become actively involved in the social and political activity of the state and thereby enhance the role of democracy. In this way both immigrants and the national citizens would be challenged to appreciate what the phenomena of immigration really turns out to be: more of a polity-shaping force and less of an economic arrangement. Add to this the fact that migrants have increasingly come to depend upon their continued residency in the country in order to lead a meaningful life, then equal civic rights are both morally and constitutionally mandated.

As the ability to decide exactly who can and cannot be a citizen has traditionally been viewed as a special sovereign attribute belonging to the state, the automatic recognition of resident migrants implies a restriction on this power of the state to define itself by freely deciding about its membership. However, the commitment to liberal democracy requires this restriction in view of the increase in human mobility. As the author puts it, '[i]t is a restriction that the commitment to democratic citizenship sets on the very definition of national citizenship' (p 20). Automatic membership does not in principle question the legitimacy of state power to decide who is a citizen but it does require some limitation on this power if the state is committed to a liberal democratic constitutionalism.

The problem of illegal immigration is dealt with in a separate chapter. It might be thought that the position of such people would differ from that of legal migrants. After all, illegal entrants have either never been allowed to enter the territory of the state or have been refused entry. Why then should the state be under an obligation to recognise automatically and unconditionally such people as citizens despite the fact that several years may have elapsed since they gained entry? Some states either tolerate illegal immigrants by allowing communities of them to build up or fail to provide sufficient resources to the removal agencies to perform their tasks effectively. Either way deportation of such illegal entrants is rendered difficult. At the same time, these people are permanently subject to the law and deeply affected by the political process. However, illegal entrants are more marginalised than legally resident migrants as they are not generally able to avail themselves of the same social benefits that legal migrants can. Therefore, it is concluded that the effects of keeping such people out of the realm of civic equality can only be more severe and therefore places them in more need of inclusion.

The author's detailed assessments of the constitutional jurisprudence of the United States and Germany provide case-studies of how two states have struggled with the issue and complement the substantive argument. While both countries have made some steps towards recognizing the constitutional relevance of the integration of migrants into society, they have not yet adopted the policy of automatic and unconditional recognition advocated by the author. The US Supreme Court therefore gets taken to task for refusing to overrule its doctrine granting the federal powers broad discretion as regards the deportation of well-settled immigrants. While the German Federal Constitutional Court has recognized resident aliens as entitled to social rights and benefits, it has refused to afford them rights to political representation and civic equality. Such rights are perceived as essentially connected to Germany's commitment to a democratic state and therefore have applied

exclusively to citizens alone. In opposing the attempts of two Länder to grant local suffrage to resident aliens, the Federal Constitutional Court has expressly rejected the path of full inclusion and the possibility of disentangling the full enjoyment of rights from citizenship. The author notes that while the future is uncertain, it will essentially depend upon two competing visions of the German state. The first vision is one in which the German state is a foundational political community where the locus of democratic sovereignty lies in national homogeneity and the loyalty of the citizen. The second vision focuses upon the state as committed to a liberal and pluralistic democracy which in turn demands a redefinition of German national citizenship laws in more inclusive terms. Overall, the author concludes that while the higher courts in both countries have taken important steps towards the recognition of the constitutional relevance of the societal integration of a non-citizen population, they have some way to go before they can offer the full inclusion and civic equality advocated here and overcome some of the old and outdated assumptions based on state sovereignty.

This book then offers a detailed constitutional argument in favour of the recognition of resident migrants as citizens. The argument is certainly extensive though at times this reviewer found it to be somewhat repetitive. The principal difficulty with the thesis of the book, however, is that there are significant practical obstacles in its implementation. Immigration and nationality are highly sensitive areas of policy and national governments are reluctant to restrict the use of their powers in such areas. While states may in practice eventually recognize long-term migrants as national citizens, they are unlikely to give up voluntarily or limit their powers as regards citizenship. The temptation to retain discretion is too powerful in political terms despite any constitutional claim that the state is informed by the values of liberal democracy. Furthermore, states are unsurprisingly reluctant to recognize illegal immigrants as citizens regardless of how long they may have been resident within their territory. Again, the political imperative of maintaining effective border controls and enforcement policies are likely to override any constitutional claims. The book then comes across as a political wish-list wrapped up in the language of constitutionalism. Nothing wrong with that, it might be said. However, the argument developed here does imply that modern states have a constitutional commitment to liberal democracy which they wish to fulfil by promoting full inclusion in the political realm of resident aliens. There is of course no guarantee that all Western states would accept these assumptions. For instance, if we seek to apply the book's argument to the United Kingdom, then the practical difficulties are compounded. The United Kingdom lacks a concept of citizenship which entails participation rights for individuals even as regards its own population. Rather, such rights are privileges enjoyed by the national subject. The decision to confer British nationality upon an individual is pre-eminently a discretionary one taken by the government and was not until recently subject to judicial review with the need for applicants to be informed of the case against them. But the argument advanced here by Rubio-Marín would be given short shrift for the more basic reason that the United Kingdom does not, despite the passage of the Human Rights Act 1998, have a constitutional commitment to upholding the values of liberal democracy.

*Robert Thomas**

* School of Law, University of Manchester.

Information for Contributors

Contributions to and correspondence concerning the various sections of the Review should be sent to the relevant Editor:

Articles: Hugh Collins, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Legislation and Reports: R.C. Simpson, Law Department, London School of Economics, Houghton Street, London WC2A 2AE

Case Notes: Colin Scott, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Books for Review, Review Articles and Reviews: Neil Duxbury, Faculty of Law, University of Manchester, Oxford Road, Manchester M13 9PL.

General Correspondence. Tim Murphy, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Section Editors will be happy to discuss drafts and proposed contributions with contributors. All contributions should be typed, double spaced and include an estimate of the number of words, including footnotes. Articles should be preceded by a short abstract of approximately 100–150 words. Reports, Legislation and Case Notes should be preceded by a list of all reports, legislation and cases cited in the paper, displayed as keywords. Contributions should use footnotes which make cross-references within the text as sparingly as possible. Contributions must conform to the Review's house style; a style sheet is available on the MLR website or will be provided by the Production Editor on request.

Articles do not usually exceed 12,000 words (excluding footnotes). Longer or shorter articles can be considered.

Subscription Information 2001

The Modern Law Review is published six times a year. Subscription rates are as follows:

Annual subscription for unbound parts only

UK and Europe	£108.00
Rest of the world	£114.00
The Americas	US\$186.00*
UK and European law teachers and law students	£25.00
Commonwealth law teachers and law students	£32.00/US\$50.00*

Annual subscriptions for Bound Volume Service: unbound parts plus a bound volume

UK and Europe	£159.00
Rest of the world	£169.00
The Americas	US\$266.00*
UK and European law teachers and students	£65.00
Commonwealth law teachers and law students	£68.00/US\$109.00*

Back Issues

Single issues of the unbound parts of the last three volumes are available from the publishers at the current price for unbound parts:

Single bound volumes	£22.00 (£23.00 for ROW, US\$37.00 The Americas)*
Complete sets of bound volumes	£71.00 (£74.00 for ROW, US\$122.00 The Americas)*
	£2071.00 (US\$3417.00 The Americas)*

*Canadian customers/residents please add 7% GST.

Orders, sample copy requests, claims and other subscription enquiries should be addressed to: Blackwell Publishers Journals, PO Box 805, 108 Cowley Road, Oxford OX4 1FH, UK. All subscriptions are supplied on a calendar year basis (January to December).

US Mailing

Periodicals postage paid at Rahway, New Jersey.

Postmaster: send address corrections to *The Modern Law Review*, c/o Mercury Airfreight International Ltd Inc., 365 Blair Road, Avenel, NJ 07001, USA (US mailing agent).

Microform

The journal is available on microfilm (16 mm or 35 mm) or 105 mm microfiche from the Serials Acquisitions Department, Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106, USA.

VOL. 61, NO. 3

MAY 2001

THE MODERN LAW REVIEW

PUBLISHED FOR THE MODERN LAW REVIEW LIMITED
BY BLACKWELL PUBLISHERS, OXFORD, UK AND BOSTON, USA

THE MODERN LAW REVIEW

Volume 64

EDITORIAL BOARD

J.S. Anderson
M.R. Chesterman
Lord Dahrendorf of Clare Market
Gerald Dworkin
J.M. Evans
Anthony Giddens
Cyril Glasser
J.A.G. Griffith

Carol Harlow
Harvey McGregor
T.M. Partington
I.G. Patel
Lord Wedderburn of Charlton, *Editor Emeritus*
Michael Zander

EDITORIAL COMMITTEE

Julia Black
Roger Brownsword
Hugh Collins, *Articles Editor*
W.R. Cornish
Neil Duxbury, *Reviews Editor*
Judith Freedman
Nicola Lacey
Martin Loughlin
Kate Maleson, *Production Editor*

W.T. Murphy, *General Editor*
R.W. Rawlings
S.A. Roberts
Colin Scott, *Case Note Editor*
R.C. Simpson, *Legislation and Reports Editor*
F.G. Snyder
Erika Szyszczak
Sarah Worthington, *Secretary*

Published for The Modern Law Review Limited by Blackwell Publishers.

© The Modern Law Review Limited 2001 ISSN 0026-7961

The Modern Law Review Ltd is incorporated in England as a company limited by guarantee under the Companies Act (Reg No 325 282) and is registered as a charity (Reg No 264072)

The Modern Law Review is published six times a year, in January, March, May, July, September and November by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main St, Malden, MA 02148, USA

Information for contributors, subscription information and details about microfilm and advertising appear on the inside back cover.

The Modern Law Review website is at <http://www.lse.ac.uk/depts/law/mlr.htm>

Printed and bound in Great Britain by The Alden Press, Oxford
This journal is printed on acid-free paper.

COPYRIGHT: All rights reserved. Apart from fair dealing for the purposes of research or private study, or criticism or review, as permitted under the UK Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing of the Publisher, or in accordance with the terms of photocopying licences issued by organisations authorised by the Publisher to administer reprographic rights, such as the Copyright Licensing Agency (CLA) and the Copyright Clearance Center (CCC). Institutions with a paid subscription to this journal may make photocopies for teaching purposes free of charge provided such copies are not resold.

CONTENTS



ARTICLES

- The Law: An Engine for Trade 333
The Right Honourable The Lord Irvine of Lairg

- In Search of the Responsible Subject: History, Philosophy and
 Social Sciences in Criminal Law Theory 350
Nicola Lacey

- Police Complaints and Criminal Prosecutions 372
Graham Smith

- Multiple Wrongdoing and Offence Structure: A Plea for Consistency and
 Fair Labelling 393
Barry Mitchell

LEGISLATION

- The Immigration and Asylum Act 1999: A Missed Opportunity? 413
Dallal Stevens

REPORTS

- Reforming Directors' Duties 439
Sarah Worthington

- Integrated Ombudsmanry: Joined-up to a Point 459
Brian Thompson

CASES

- What Price Auctions Without Reserve? 468
Frank Meisel

- One Immunity Has Gone ... Another ...: *Holland v Lampen-Wolfe* 472
Mizushima Tomonori

- Sale and Lease-Back Agreements in a World of Title Relativity:
Michael Gerson (Leasing) Ltd v Wilkinson and State Securities Ltd 481
Janet Ulph

- Only a Matter of Time 489
Nicole Busby

REVIEW ARTICLE

- The Theoretical Controversy Concerning Judicial Review 500
Andrew Halpin

REVIEWS

- Maclean: Making Law for Families 512
 Kahn: The Cultural Study of Law: Reconstructing Legal Scholarship 515
 Ahdar: Law and Religion 516

Modern Law Review Seminars

The *Modern Law Review* has decided to make available up to £25,000 for 2002 to fund the organisation of one or more scholarly seminars on any subject broadly within the publishing interests of the *Review*.

Applications for funding should be submitted to the Secretary of the *Review* no later than 12 October 2001.

Guidelines on application for awards are available from Sarah Worthington, Modern Law Review, Law Department, London School of Economics, Houghton Street, London, WC2A 2AE. Email s.worthington@lse.ac.uk. Additionally the Guidelines are published on the *Modern Law Review* Home Page at <http://www.lse.ac.uk/depts/law/mlr.htm>

THE MODERN LAW REVIEW

Volume 64

May 2001

No 3

The Law: An Engine for Trade

*The Right Honourable The Lord Irvine of Lairg**

The lecture traces the process of commercialisation in English law from its early stages to the present day. Until the mid-eighteenth century the law was in a process of integration, overcoming a judicial preoccupation with technicalities and procedures to form a body of rules which merchants could trust. From that point on the law has assumed a proactive role as an engine for trade. The nineteenth century is marked by legal facilitation, where new institutions were fashioned as a robust commercial framework for the Industrial Revolution. From the end of that century to the present day business law has served to regulate, building a sensitive framework for commercial development that balances the needs of commerce with the needs of society as a whole.

My subject is commercial law and its role as an engine for trade. English law is practically unique among the legal systems of the world because its commercial rules are entirely integrated into the general practice of the law, with business disputes heard in the same courts, using the same principles, as other litigation.

This sets commercial principles in context. It avoids stark distinctions between commercial contracts, private contracts and consumer contracts and stands in the way of overclassification. The explanation for separate commercial systems elsewhere is that they should reduce legal burdens on business, whilst an integrated system may not provide the fast and sensitive procedures traders require.¹

English law, however, is highly regarded in the international commercial community. Out of seventy-two trials heard in the Commercial List in the High Court during the last year, forty-four involved foreign parties.

Let us consider the four principal ways in which commercial law serves our free market economy. First, commercial law is clear and predictable, providing a firm body of rules on which traders can depend. Secondly, it contains a strong and positive law of contract to uphold trading agreements. Thirdly, the law has been shaped by the needs and expectations of merchants, and much business practice has been incorporated into law. Fourthly, Parliament and the courts have made genuine innovations to lead the market economy forward, building on business practice to provide strong frameworks for industry.

* A lecture delivered to the British Academy on Wednesday 22 November 2000. I acknowledge my indebtedness to Neil Beraford, Barrister-at-law, for his invaluable assistance in the preparation of this lecture.

¹ In France, for example, commercial contracts are an exception to the strict rule, contained in Civil Code, Art 1341, that contracts must be proved by signed documents. Art 109 of the Commercial Code allows commercial contracts to be evidenced *par tous moyens*. The limitation period is also shorter (being 10 years) than the 30-year period applicable to general contractual transactions: Commercial Code Art 189.

Legal predictability

The very first need of the business community is legal predictability. An unpredictable legal climate is unacceptable to business, forcing traders into necessary legal advice and insurance cover to secure against the risk of their deals being defeated. In 1774, Lord Mansfield said:

In all mercantile transactions the great object should be certainty; and therefore, it is of more consequence that a rule should be certain, than whether the rule is established one way or the other. Because speculators in trade then know what ground to go upon.

The commercialisation of the common law

Before 1750, commerce was hindered by a division between the law merchant and the common law.² Common law was the unified law of the land,³ yet it was slow to innovate in commerce, since it had become preoccupied with its own technicalities and procedures rather than the requirements of business.⁴ A misplaced word in a property transfer would often defeat entire transactions.⁵ Businessmen had to have some reprieve from these stern rules, and the solution was to let merchants rely on their own customs as exceptions to the common law. If a transaction accorded with a practice recognised in the mercantile community, its validity could be assessed by reference to that practice rather than the insensitive common law.⁶

The solution was an obvious compromise, the price of which was legal certainty.⁷ Litigating businessmen invariably put evidence of local customs before their judge,⁸ and the success or failure of commercial transactions then came to depend on the judge's acceptance of this factual evidence.⁹

It was not until Lord Mansfield became Lord Chief Justice in 1756 that positive measures were taken to combat the uncertainty permeating commercial litigation. Mansfield's approach was to incorporate mercantile customs directly into the common law. When a merchant led evidence of a local custom, Mansfield evaluated its content by consulting businessmen¹⁰ and applying his renowned expertise in foreign systems.¹¹ If the custom was accepted, it would become

2 Maitland, 'Select Pleas in Manorial Courts' (1889) 2 *Selden Soc* 132.

3 J. Holden, *The History of Negotiable Instruments in English Law* (London: University of London, Athlone Press, 1955) 27–36, M. Postan, *Medieval Trade and Finance* (Cambridge: Cambridge University Press, 1973) ch 2. For a different perspective on the nature of *lex mercatoria* (but a similar conclusion on the question of incorporation) see J. Baker, 'The Law Merchant and the Common Law Before 1700' [1979] *CLJ* 295.

4 P.S. Atiyah, *The Rise and Fall of Freedom of Contract* (Oxford: Oxford University Press, 1979) ch 12.

5 J. Powell, *Essay upon the Law of Contracts and Agreements* (1790) I, 152–160.

6 If he did not adduce such evidence, the merchant would be 'stuck' in the common law doctrine of *assumpsit*, and the custom would only be relevant to 'explain the *assumpsit*', that is to justify the implication of a liability: *Oaste v Taylor* (1612) *Cro Jac* 306.

7 See the complaints of Buller J in *Luckbarrow v Mason* (1793) 2 *TR* 73.

8 This in fact appears to be what the majority of business litigants did: *Death v Serwonters* (1685) *Lutw* 885.

9 It was eventually settled that merchants wishing to rely on trade customs were obliged to meet two heavy burdens of proof: that the custom had immemorial antiquity and limited geographical application: *Brown v London* (1669) 1 *Lev* 298.

10 See, for example, his evaluation of the rules of contract in *Loveacres d Mudge v Blight* (1775) 1 *Cowp* 352.

11 In the leading case of *Lewis v Rucker* (1761) 2 *Burr* 1167, Lord Mansfield determined the method of quantifying an insurer's liability for partial loss by first 'conversing with some gentlemen of experience in adjustments'.

binding law,¹² and no subsequent court could admit further evidence on the point.¹³

Through this process, Lord Mansfield laid the foundation of the modern commercial system. Commercial law was transformed from propositions of fact into a rational corpus of law, capable of consultation by businessmen, lawyers and judges alike.¹⁴

Late nineteenth-century codification projects

In the nineteenth century the exponential growth in commerce from the Industrial Revolution fuelled the demand for clear and accessible commercial rules. Digests of cases on core commercial subjects were perceived as inadequate when traders looked to the commercial codes of Continental Europe and India,¹⁵ and the Associated Chambers of Commerce seized the initiative.¹⁶ During the 1880s, leading commercial minds of the day¹⁷ were commissioned to collate the case law into a cluster of codifying legislation in the four key areas of commercial activity: bills of exchange, factoring, partnerships and sales of goods.¹⁸

The codes were a huge advance in the quest for legal clarity. Adopted throughout the Commonwealth, they condensed a hundred years of common law understanding into individual statutory statements. Their format relieved traders of the need to refer to the vast body of case law when ascertaining commercial rules. In *Bank of England v Vagliano Brothers*,¹⁹ concerning the Bills of Exchange Act 1882, Lord Herschell said:²⁰

The purpose of [a codifying] statute surely was that on any point specifically dealt with by it, the law should be ascertained by interpreting the language used instead of, as before, by roaming over a vast number of authorities in order to discover what the law was, extracting it by a minute critical examination of the prior decisions.

The statutes were a triumph, described by one judge as 'the best-drafted Acts of Parliament ever passed'.²¹ Not only did they make the law accessible,²² but they yielded such effective principles that the law has required little reform since.²³

12 *Edie v East India Co* (1761) 2 Burr 1216.

13 Evidence in future would thus be admissible only on matters which had not yet been resolved: *Long v Allen* (1785), detailed in J. Park, *Marine Insurance* (1st ed) 446.

14 Commentaries were soon produced to rationalise particular areas. Among the earliest commentators were Paley, *Principles of Moral and Political Philosophy* (1785), see Powell, n 5 above. These were soon followed by famous works on more specific subjects, such as Chitty's *Treatise on the Law of Bills of Exchange* (1799) and Byles' *Treatise on the Law of Bills of Exchange* (1829).

15 M. Chalmers, 'An Experiment in Codification' (1886) 2 LQR 125; J. Dove-Wilson, 'Concerning a Code of Commercial Law' (1884) 28 *Journal of Jurisprudence* 337.

16 For the role of business in instigating the codification process, see B. Rodger, 'The Codification of Commercial Law in Victorian Britain' (1992) 108 LQR 570.

17 Sir Frederick Pollock, author of a highly influential digest on the law of partnership, drafted the Partnership Bill. The rest were entrusted in large part to Sir Mackenzie Chalmers.

18 The cost of the draft Bills was ultimately shared between the Chambers of Commerce and the Institute of Bankers: Executive Council of the Associated Chambers of Commerce, 11 March 1881.

19 [1891] AC 107.

20 At 144-145.

21 MacKinnon, LJ in *Bank Polski v Mulder & Co* [1942] 1 All ER 396.

22 Another argument in favour of codification was that it would offer a single coherent statement of English business law at a time when numerous international conventions were being held to undertake the global harmonisation of trade laws: T. Barclay, 'Assimilation of Mercantile Law' (1886) 2 LQR 66.

23 The codes also gave the opportunity for the assimilation of laws in England and Scotland. These laws remained very different even following the investigation of a Royal Commission in June 1853. The

This is not to say that the law has stood still. Many details have required modification to reflect changing commercial and social values, but this modification has taken place well within the original framework. In sales law for example, demand from consumers has tightened the legal duties imposed on sellers,²⁴ and demand from business has necessitated the legal recognition of co-ownership in mixed goods.²⁵ Parliament has introduced these changes without compromising the clarity of the original code, by passing consolidating legislation. The Sale of Goods Act 1979 brings together the disparate Sales Acts passed since 1893, reproducing the original Act and its subsequent reforms in a single instrument. This consolidation does not serve the same function as codification, since earlier materials are not replaced, merely republished. However, the combination of both measures – codification and consolidation – has ensured that our commercial laws remain both relevant and accessible.

Parliament has even added entire branches of law without disturbing the integrity of the original legislation. This summer Parliament passed the most important partnerships legislation since the original Act of 1890 in the shape of the Limited Liability Partnerships Act.²⁶ The Act will add another form of business entity to the existing armoury, providing an attractive means of incorporation for multinational accountancy and legal firms.

Nineteenth-century codification was therefore a great success, which has led some to question why the process should not be expanded across a broader ambit of commercial transactions.²⁷ The reason for this lies in the law's inherent pragmatism. During the nineteenth-century projects, full codification was considered to be too difficult within the British parliamentary system, and lawyers concluded that the complexity of formulating and enacting a large commercial code would not be justified by positive benefits.²⁸ Experience from France and Germany has added little momentum to the campaign, as the commercial codes in these countries have yielded only equivocal benefits. No code can be entirely comprehensive,²⁹ and in Germany a trader now refers to the commercial code for the content of his agreements, to the civil code for the existence of his agreements, and to judicial precedent for the validity of his agreements.³⁰

Commission published a first Report in 1854, and legislation followed in 1856. However, it was unpopular and widely criticised at the time for making unsatisfactory and superfluous changes to the law. 'English Amendments of Scotch Law', 1 Scottish LJ 1 (1858)

24 The most significant amendments were made by the Supply of Goods (Implied Terms) Act 1973, which introduced a distinction between consumer and non-consumer transactions and introduced an implied term of merchantable (now *satisfactory*) quality where the seller is in the course of a business. It also took measures to prevent sellers from restricting their liability and consolidated the terms implied into contracts of sale with those implied into contracts of hire purchase. For details, see I. Carr, 'The Supply of Goods (Implied Terms) Act 1973' (1973) 36 MLR 519

25 See the Sale of Goods (Amendment) Act 1995.

26 See N. Beresford 149 (1999) NLJ 1647

27 See, eg R. Goode, *Commercial Law in the Next Millennium* (London: Sweet & Maxwell, 1998) ch 4.

28 It was said by no less a reformer than Lord Selborne LC, chief proponent of the Judicature Acts, that a wholesale commercial code would create enormous difficulties in Parliament, and would be unlikely to pass through both Houses without so much scrutiny as to destroy the utility of the project: 'Codification of Commercial Law' (1885) 78 *Law Times* 321.

29 The German commentator Kondgen writes of the 1897 *Handelsgesetzbuch* (Commercial Code) 'Its choice of transaction types seems selective at best, arbitrary at worst; and its provisions are often fragmentary. While we find adequate provisions relating to the carriage and storage of goods, or commercial agency, the Code is silent on banking, secured transactions or commercial leases; the provisions on commercial sales cover merely ten sections.' ('Commercial Law', in W. Ebke and M. Pinkin (eds), *Introduction to German Law* (The Hague/London: Kluwer, 1996))

30 Under the principle of good faith, devised by German courts in response to hyperinflation during the 1920s and 1930s and based upon § 242 BGB.

In Britain, the absence of full-scale codification has brought a significant advantage: legal flexibility. Since it is founded on general principles, commercial law has evolved in tandem with wider legal values. For example, in the Victorian era from which the Sale of Goods Act dates, a contract formed under the influence of illegitimate pressure would only be set aside if this pressure took the form of a physical threat.³¹ Today we recognise a subtle body of rules that regards all illegitimate threats as capable of amounting to duress.³² By a process of steady evolution, the law has moved far from its unforgiving Victorian roots.

Laying the foundations for a free market economy: the law of contract³³

The second keystone of commercial law is its recognition of private agreements. Today it goes without saying that most types of business contract may be freely made and enforced in the courts, with the obvious exception of contracts which have unlawful or immoral purposes.³⁴ In any free market system, a large proportion of wealth is concentrated in speculative interests held on share and futures exchanges. For this market to operate, it is essential that traders be kept to their promises.

The absence of any law of contract at the turn of the nineteenth century

Yet the modern law of agreements is a surprisingly recent phenomenon. In the wake of the Glorious Revolution, secure property rights assumed supreme political importance³⁵ and legal policy was focused on the protection of ownership.³⁶ Any financial speculation undermining this security was frowned upon, and even criminalised. As late as 1800, traders were being prosecuted for offences such as *regrating*, buying goods for resale at the same market.³⁷

This economic climate had no need for an autonomous law of contract. Agreements did little more than facilitate property transfer,³⁸ and the law gave no support to market speculation. Agreements to purchase goods were unenforceable until the goods had in fact been delivered,³⁹ so parties could simply escape from a bargain if a change in the market price made it unprofitable. Even when goods had been delivered, a supplier could not insist on the agreed contract price, but

31 *Skeate v Beale* (1841) 11 A & E 983.

32 *The Dimski Shipping Corporation SA v International Transport Federation (The Evla Luck)* [1992] 2 AC 152. The law is, of course, commercially sensitive. Careful distinctions are drawn between those threats which are illegitimate and those which are legitimate business practices: *CTN Cash & Carry v Gallagher Ltd* [1994] 4 All ER 714.

33 See generally n 4 above; A.W.B. Simpson, 'Innovation in 19th Century Contract Law' (1975) 91 LQR 247.

34 See *Lipkin-Gorman v Karpnale* [1991] 2 AC 548.

35 J. Plumb, *The Growth of Political Stability in England 1675–1725* (London: Macmillan, 1967).

36 As Thompson points out: 'Millers and – to a greater degree – bakers were considered as servants of the community, working not for a profit but for a fair allowance': E. Thompson, 'The Moral Economy of the English Crowd in the Eighteenth Century' (1971) 50 *Past & Present* 76, 83.

37 *R v Ruby* (1800) Peake Add Cas 189.

38 It is of particular interest that in Blackstone's *Commentaries*, contract is dealt with under *Volume ii* ('Rights of Things').

39 In *Walker v Moore* (1829) 10 B & C 416, Littledale J. went so far as to say: 'It is contrary to the policy of the law that a man should offer an estate for sale before he has obtained possession of it'.

damages in any court proceedings would be confined to the objective value of the goods provided.⁴⁰

Nineteenth-century revolutions – the age of freedom of contract

A developed law of contract did not really begin until the 1820s, led on by the pressures of railway capitalism.⁴¹ Steam railways demanded an unprecedented financial outlay which could only be met through public subscription, and they offered healthy profits in return. Speculation took off on an unprecedented scale,⁴² but the framework for this speculation was simply not in place. With no adequate legal backing, the market experienced three devastating crashes in the space of twenty years.⁴³

In response to the vast increase in trade during the Industrial Revolution, a new and principled law of agreements emerged.⁴⁴ Its rules were focused on the original intentions of the parties to the agreement, so bringing three crucial advantages to business. First, technical and artificial constructs gave way to new simple rules for the interpretation of agreements.⁴⁵ Secondly, the law began to enforce speculative bargains without requiring prior performance by either party, preventing recalcitrant traders from evading unprofitable transactions.⁴⁶ Thirdly, damages began to be assessed on a realistic basis, the court looking to the terms of the original agreement to calculate a party's loss.⁴⁷

In this way the law has evolved into a fine engine for financial speculation. By 1877, less than eighty years after Kenyon LC had condemned regrating as contrary to the common law,⁴⁸ Bacon VC said in the case of *Noble v Edwards*,⁴⁹

A man who speculates in land means always to get as much profit from it as he can. If, by his superior skill, he foresees that he can make an advantageous profit by working, cultivating and improving a farm, certainly if it is worth his while he can do that, and it would be quite worth the while of anybody buying from him to pay him whatever in their respective judgments, the land is worth, without considering what the then vendor gave for it himself. What is more common? It is everyday practice in this court.

40 *Flureau v Thornhill* (1775) 2 W BL 1078. The reason for this is that promises were enforced not for the subjective reason that the parties had reached agreement, but only in so far as the transaction had a binding moral force.

41 See generally R. Kostal, *Law & English Railway Capitalism 1825–1875* (Oxford: Clarendon Press, 1994).

42 N. McKendrick, J. Brewer and J. Plumb, *The Birth of a Consumer Society: The Commercialisation of Eighteenth-Century England* (London: Europa, 1982).

43 The crashes took place in 1826, 1836 and 1845. Contemporary businessmen pilloried the legal system for its shortcomings. The harsh verdict of the *Railway Gazette* (26 May 1849) was that 'the public have waited in vain for three years for some speedy means of bringing to justice those who robbed them. ... They have hitherto been helplessly and hopelessly without redress.'

44 Such work was founded on the work of Lord Mansfield, who half a century earlier had championed the intentions of the parties as a means of contractual interpretation in cases such as *Pillans & Rose v Van Mierop & Hopkins* (1765) 3 Burr 1663.

45 See *Kingston v Preston* (1773) 2 Doug 691.

46 Simpson, n 33 above.

47 This was achieved by the time of *Hadley v Baxendale* (1854) 9 Ex 341.

48 *R v Ruby* (1800) Peake Add Cas 189.

49 (1877) 5 ChD 378.

Upholding the expectations of merchants – legalising mercantile practice

The law of contract gives traders the flexibility to establish business arrangements tailored to their own requirements. But a business agreement will rarely make express provision for every detail of the bargain, and so the law must fill the gaps, giving effect to the presumptions and expectations of the traders entering the agreement.

English law achieves this objective in two ways: by creating legal rules based on realistic trade customs, and by recognising that traders often wish to incorporate standard terms and practices into their agreements.

Conferring positive legal status on mercantile institutions

Let me turn to three specific illustrations of how the law has facilitated trade by creating bodies of law based on trade custom.

Bills of exchange

First, bills of exchange. A debtor wishing to make a payment may approach his bank for a bill of exchange, to be drawn in favour of the creditor. The bill is sent to the creditor, who has two options: either he may present the bill to the bank or its local agent for payment, or he may transfer the bill to a third party, who then receives the right to payment from the bank.

The institution holds many advantages for business, and has been operated by merchants since the thirteenth century. Its greatest attraction is the attribute of free transfer, or *negotiability*. The creditor receives a promise of payment that can readily be transferred to third parties with the minimum of formality.

Bills of exchange were not easy to integrate into the law. Their negotiability sits in stark contrast to the inflexible common law rule that contractual rights may not be transferred to third parties.⁵⁰ For this reason, early decisions set bills of exchange outside the common law, in the law merchant.⁵¹ By the late seventeenth century, however, bills of exchange had become such a central feature of commercial transactions that they were finally recognised by the common law,⁵² and a full set of rules was swift to develop. These rules were modelled firmly upon the practice of traders.

For example, in a case of 1695⁵³ the court was asked to rule upon the appropriate rule for the settlement of foreign bills. Evidence was led, proving that universal practice in the business world was to allow three days' grace for payment. This practice was accepted as law.

Much further development and consolidation took place during the eighteenth and nineteenth centuries – the numerous digests on the law⁵⁴ culminating in Sir Mackenzie Chalmers' Bills of Exchange Act 1882. Its principles derive directly from the practice of businessmen and that is the reason for its enduring success.

⁵⁰ In *Three Rivers v Bank of England* [1996] QB 92, the court summed up the common law's refusal to countenance the assignability of contractual rights. Assignment is only recognised in equity or in exceptional cases by statute. See Law of Property Act 1925, s 136.

⁵¹ See *Burton v Davy* (1437) SS 49, 117.

⁵² This achievement is credited to *Woodward v Rowe* (1666) 2 Keb 105, 132.

⁵³ *Tassell and Lee v Lewis* (1695) 1 Ld Raym 743.

⁵⁴ eg Chitty's *Treatise on the Law of Bills of Exchange* (1799), Byles' *Treatise on the Law of Bills of Exchange* (1829).

Bills of lading

The innovation of the international shipping community provided another fertile source of inspiration. Of particular importance to modern shipping is the bill of lading, a document issued by the carrier both to evidence the shipping contract and to prove the shipper's title to the cargo. Yet bills of lading are another legal oddity. Once issued by the carrier they may be freely conveyed to third parties, and this conveyance operates to transfer not only title to the cargo, but also the rights of the shipper under the contract of carriage.⁵⁵

Given the conflict between bills of lading and the common law's refusal to sanction transferable contracts, these bills were slow to be recognised. It took the decision of a special jury in the historic decision of *Lickbarrow v Mason*⁵⁶ in 1793 to establish that transfer of title to the cargo accompanied transfer of the bill and that became the common law. However, it took the Bills of Lading Act, 1855, to provide that the rights of the shipper under the contract of carriage also accompanied transfer of the bill.

Two unusual features highlight the unique commercial origins of the law governing bills of lading. First, legal intervention has been kept to a bare minimum, to the extent that many central doctrines remain as presumptions, not rules of law. Indeed, the first principle of shipping law – that carriers may always be compelled to issue a bill of lading – stands entirely unsupported by legal authority.⁵⁷ The law has intervened only as required, as for example in 1992, when the Carriage of Goods by Sea Act was passed to close a serious lacuna identified in a case before the House of Lords when a carrier escaped all liability for negligently damaging a cargo sold in transit.⁵⁸

The second unique feature of shipping law is the extent of international harmonisation. At the turn of the twentieth century, international practice began to grow disparate. Individual states had introduced legislation regulating the rights and duties of parties to shipping agreements,⁵⁹ but this piecemeal approach could not be sustained in an increasingly global marketplace. The international trading community moved for standardisation, and a body of international rules was introduced at The Hague on 25 August 1924.⁶⁰ The Hague Rules (as amended in 1968)⁶¹ have been highly successful in establishing a worldwide framework for contracts of carriage by sea. The rules stand as testament to the role of the law as an engine for trade. By harmonising conditions of carriage across the globe, they have lowered the costs of negotiating international contracts and established minimum standards to set a level playing field for traders.

55 Since the Carriage of Goods by Sea Act 1992.

56 London: Shaw, (1793) 2 TR 73.

57 L. D'Arcy, *Ridley's Law of the Carriage of Goods by Land, Sea and Air* (London: Shaw, 7th ed, 1992) 82.

58 See *Leigh & Silavan v The Allakmon Shipping Company Ltd, The Allakmon* [1986] AC 785, where the buyer of steel coils was left without a remedy when the goods were negligently damaged in transit at a time when they still belonged to the sellers. The sellers had no remedy as they had suffered no loss, and the contractual rights under the agreement had not passed to the buyers until a point in time after the damage had already occurred.

59 The first legislation of the modern age was the Harter Act, enacted in the United States in 1893. This was followed by the Australian Sea Carriage of Goods Act 1904 and the 1910 Water Carriage of Goods Act in Canada.

60 The International Convention for the Unification of Certain Rules of Law Relating to Bills of Lading, 25 August 1924. The Convention was implemented into English law by the Carriage of Goods by Sea Act 1924, which came into force on 1 January 1925.

61 This amendment was necessary to take account of the container revolution. It was signed at Brussels on 23 February 1968, and implemented into English law by the Carriage of Goods by Sea Act 1971.

Letters of credit

My third example of business-led development in commercial law comes from international finance. Letters of credit are financing mechanisms operated between parties to an international sale. The buyer opens at his bank a credit in favour of the seller subject to specific instructions, usually that the bank withholds payment until it receives valid shipping documents. The buyer's bank then sends a letter of credit to the seller, either through its own offices or through another bank in the seller's country. Once the goods have been dispatched, the seller must present the requisite documents to the bank, and payment will be made.

Letters of credit provide traders in the international market with a guaranteed means of payment, which oils the flow of international trade. Yet letters of credit contain two legal surprises. First, they are anomalous within the common law. By mercantile usage, the bank's payment undertaking is considered binding by virtue of its issue alone without the need for consideration or reliance by the beneficiary. A second surprise is the complete absence of statutory intervention in this area. Both these features are understandable, however, since letters of credit have developed almost entirely through the initiative of businessmen.

In 1933, the International Chamber of Commerce published a set of Uniform Customs and Practice for Documentary Credits to establish an international standard for letters of credit. Revised on a regular basis,⁶² the Uniform Customs and Practice are incorporated into letters of credit worldwide.⁶³ They are a prime example of commercial development taking place outside the legal arena, but recognised by the law. Of course, letters of credit have come before the courts and judges have upheld their legality as an aspect of mercantile practice, but these judgments owe their content largely to the terms of the Uniform Customs and Practice. In 1983, the House of Lords was asked to determine a point on which there was no prior authority in English law: whether, when the seller had unwittingly tendered to the confirming bank documents containing a false statement, the bank was obliged to pay.⁶⁴ The House of Lords found that the bank *was* obliged to pay, after evaluating the possible solutions to the problem with one eye to the Uniform Customs and Practice and another to the demands of commerce. Lord Diplock dismissed the proposition, that the bank was not obliged to pay if the documents, although conforming on their face with the terms of the credit, contained some inaccurate statement of material fact, thus:⁶⁵

The more closely this bold proposition is subjected to legal analysis, the more implausible it becomes; to assent to it would, in my view, undermine the whole system of financing international trade by means of documentary credits.

Letters of credit are thus another prime example of the law responding to change by adapting its rules with keen commercial sensitivity to the customs and expectations of traders.

⁶² Revisions occurred in 1951, 1962, 1974, 1983 and 1993.

⁶³ See the comments of Mustill LJ in *Royal Bank of Scotland v Cassa di Risparmio delle Provincie Lombarde* [1992] 1 Bank LR 251 at 256, and for a full discussion, see B. Kozolchik, 'Letters of Credit' 9 *International Encyclopaedia of Comparative Law*.

⁶⁴ *United City Merchants v Royal Bank of Canada* [1983] 1 AC 168.

⁶⁵ At 184.

Soft law – incorporating the understandings of merchants into commercial agreements

The second way in which the law has acted to uphold the expectations of traders is by giving effect to standard business practices in commercial agreements. Widely accepted practices are of great value to business, as they provide ready-made bodies of rules that adapt to change and avoid the need for direct legislation. This is particularly true in international trade, where the commercial pressure for legal harmonisation is great but the potential for agreement among governments is low. The legal vacuum remains to be filled by the initiative of businessmen.

At an international level, much positive work has been carried out by the International Chamber of Commerce, (the ICC), an organisation run by businessmen from over 130 member states. We have already seen the value of the Uniform Customs and Practice for Documentary Credits. The ICC has also been active in international shipping, by drawing up a set of standard terms for international shipping contracts. INCOTERMS are drafted by lawyers and businessmen and updated regularly.⁶⁶ They are so widely accepted throughout the world that INCOTERMS are likely to be implied into shipping contracts even if the parties make no direct election themselves.

Such permissive forms of regulation are also evident in domestic trade law. The Companies Act 1985 contains a model constitution for limited liability companies, whose provisions may be adopted or excluded at will. These measures are a highly efficient way of supervising business, as they give entrepreneurs commercial freedom to select the most appropriate rules for their individual needs.⁶⁷

So, we have seen various means in which the law has facilitated trade by giving full effect to business customs. In some instances the law has followed the initiative of business by turning commercial practice into law. In others it has maintained a deliberately low profile, allowing businessmen to make their own rules and giving legal effect to these rules only when required.

Sensitive regulation of trade

Next, let me turn to instances where lawyers have led the way.

Early political economists declared that the sole objective of business law was to follow the lead of merchants. In his *Wealth of Nations*, Adam Smith concluded that man's pursuit of his own self-interest provided the key to prosperity, and that legal intervention was unwelcome in the free market process.⁶⁸ Of course, the idea has much truth. Healthy market freedom is an important interest, and modern economists continue to stress the importance of minimum commercial regulation.⁶⁹

However, at some point regulation becomes inevitable. At the time of the first railway company flotations in the 1820s, there was an entirely unregulated share

⁶⁶ The latest version is INCOTERMS 2000.

⁶⁷ B. Cheffins, *Company Law. Theory, Structure and Operation*, (Oxford: Oxford University Press, 1997), ch 5.

⁶⁸ G. Brennan and J. Buchanan, *The Power to Tax: Analytical Foundations of a Fiscal Constitution* (Cambridge: Cambridge University Press, 1980).

⁶⁹ As Professor Cheffins writes in his economic analysis of company regulation, 'The enactment and amendment of legal standards can be a process plagued by delays. ... The individuals who promulgate and administer laws governing market behaviour may not have sufficient expertise concerning the conduct being regulated to judge accurately the impact of the decisions they are taking. ... Government regulation can suffer from problems which can undermine the case for intervention': n 67 above, 364.

market.⁷⁰ Prevented from forming public companies by the notorious Bubble Act of 1721,⁷¹ entrepreneurs offered shares to the public unlawfully by trading under 'deed of settlement' associations. Disreputable enterprises flourished. Of the 624 associations floated in 1824, only one fifth survived until 1827.⁷² When the practice of public subscription was legalised in 1825, businessmen were able to operate within the law but little overall benefit was achieved. In the absence of any regulatory framework, commercial standards remained reprehensibly low. Dishonest businessmen floated thinly-capitalised companies,⁷³ and devastating stock market crashes occurred in 1826, 1836 and 1846.⁷⁴

This experience highlights the first danger of an unregulated market: the inevitable temptation to perpetrate fraud.

The second danger is the so-called externalities problem.⁷⁵ Businessmen are likely to operate to maximise their own profits, although this may prejudice society as a whole. Thus, law must regulate to ensure that the driving force of commerce does not undermine wider public interests.

Self-regulation is one possible solution to these two dangers, allowing business to take collective notice of public policy. For this reason, self-regulation is an essential feature of the English business system, and is directly encouraged by law. The Financial Services Act 1986 established a comprehensive scheme of self-regulation in the supervision of equity markets,⁷⁶ with daily operations overseen by Recognised Investment Exchanges, run by market practitioners, with ultimate authority in the Securities and Investment Board, itself a private company. The most recent Statute is the Financial Services and Markets Act 2000, which established the Financial Services Authority.⁷⁷

However, self-regulation is not always adequate, and the law must often take a positive lead in prescribing commercial standards. In taking this lead, the law recognises three key principles. First, the principle of minimum regulation: the regulator must avoid tying the hands of business with an excessive number of rules. Secondly, the principle of appropriate regulation: selecting the correct type of rule for the context. Broad mandatory rules are rarely required, and it is often possible to regulate in less intrusive ways by providing standard practices in which industry is strongly encouraged to participate. Thirdly, there is the principle of continuing relevance: the regulator must ensure that the rules passed are capable of frequent and effective updating to take account of commercial developments.

By respecting these principles in all forms of regulation, English law has been consistently sensitive to commercial needs.

70 A. B. DuBois, *The English Business Company After the Bubble Act* (London: H. Milford, OUP, 1938).

71 See generally P. L. Davies, *Gower's Principles of Modern Company Law* (London: Sweet & Maxwell, 1997), chs 2-3.

72 H. English, *A Complete View of Joint Stock Companies Formed in 1824 and 1825* (London: Sweet & Maxwell, 1827).

73 B. C. Hunt, *The Development of the Business Corporation in England 1800-1867* (New York: Russell, 1969).

74 M. Reed, *Investment in Railways in Britain 1820-1844: A Study in the Development of the Capital Market* (Oxford: Oxford University Press, 1975), 14-16; Pollins, 'The Marketing of Railway Shares in the First Half of the Nineteenth Century' (1954-55) 7 *English Historical Review* 233.

75 As identified in the ground-breaking article by Professor Ronald Coase, 'The Problem of Social Cost' (1960) 3 *Journal of Law & Economics* 1.

76 R. Bagott, 'Regulatory Reform in Britain: The Changing Face of Self-Regulation' (1989) 67 *Public Administration* 435.

77 The regulation of markets and financial services is now supervised by the Financial Services Authority under a new statutory framework created by the Financial Services and Markets Act 2000. The Act lays down four central objectives to be respected in the regulation of financial markets: market confidence, public awareness, the protection of consumers, and the reduction of financial crime.

Company law – sensitivity in Parliament and the courts

The boom-and-bust experience of railway capitalism illustrated the need for state intervention in the companies' market. Yet, at the same time, a middle course was required between total deregulation and the stifling administrative requirements formerly imposed by the Bubble Act.⁷⁸ In the aftermath of the appalling market crashes came the modern corpus of company law.

The answer to companies regulation was devised by Gladstone in 1844.⁷⁹ Since that time, the company as a business vehicle has taken off as an unequivocal success. In 1885, there were 60 listed companies. Today, the total number of companies registered at Companies House is around 1.14 million,⁸⁰ and 2,450 of these are listed on the London Stock Exchange.

The modern system rests on three key principles: free incorporation, limited liability and minimum external interference.

The first principle is that businessmen must have freedom to seek access to public money by incorporating as companies. There is of course the danger that incorporated companies will be used as a sham to swindle investors, but this danger cannot be addressed by burdensome regulation, or else promoters will operate outside the law as they did in the 1820s. The solution is a model of free market theory, a system of public company registration which is convenient for businessmen, yet informative for potential investors. Individuals now have the right to incorporate as companies, but the price of this right is the publication of details including the company's constitutional documents and annual accounts at Companies House.

The second principle is that the public must be actively encouraged to subscribe in business ventures, which generally requires that they be protected from the risk of business failure. In 1855, the Companies Act was amended to allow the liability of members to be limited to the value of their shareholdings. Again, however, there must be a price for this broad right, and Parliament has evolved protective mechanisms to ensure that limited liability is not abused at the expense of company creditors. For example, it is a rule of fundamental importance that the share capital of a company – that proportion of assets which has been contributed by shareholders – must be kept intact for the benefit of creditors. The Companies Act strictly controls the circumstances in which a company can reduce its share capital, and court approval must be sought before making a formal reduction of the capital fund.⁸¹ The common law has also evolved doctrines to prevent fraud, and shareholders will be personally liable for a company's debts if they have formed the company to perpetrate a fraud.⁸²

Thirdly there is the principle of minimum interference in company management. Informed market freedom underlies the operation of English companies legislation, and supervision of company management is generally left to the shareholders

78 As the *Railway Gazette* recorded in 1859: 'Each separate railway company which has been under the necessity of applying to Parliament for powers ... has been mulcted in an enormous amount of money for no earthly purpose but to fill the pockets of parliamentary lawyers.' *Railway Gazette*, 5 March 1859.

79 *Parliamentary Papers*, Commons (1844) VII Q 2054. See also F. Hyde, *Mr Gladstone at the Board of Trade* (1934).

80 Figures taken from DTI Consultation Paper: *Modern Company Law for a Competitive Economy* (1998).

81 This was achieved by a legislative reform of 1928. See Company Law Committee, *Section 151 Companies Act 1985* (Law Society Legal Practice Directorate, 1990) 21–32.

82 *Williams v Natural Life Health Foods Ltd* [1997] 1 BCLC 131.

acting through the general meeting. Of course, there has been some legislative intervention to provide for cases where the principle of informed market freedom has failed. Under the Company Directors Disqualification Act 1986, dishonest or incompetent company directors can be disqualified from company management. The Insolvency Act 1986 contains provisions to hold directors criminally liable for fraudulent trading,⁸³ and liable at civil law for trading in the knowledge that the company cannot pay its debts.⁸⁴ The Law Commission has recently recommended that the law impose a strict objective standard of care on company directors to ensure the better protection of those dealing with companies.⁸⁵ This would be a considerable development from the position at the turn of the twentieth century, which merely required a company director to exercise diligence in the conduct of his duties.⁸⁶

Companies regulation should not only be effective, but it must also be clear. The original Companies Act 1844 has now been amended on many separate occasions, and while these amendments are presently consolidated in the Companies Act 1985, the profusion of company law sources remains a cause for some concern. A major review of company law is currently underway at the Department of Trade and Industry to streamline the regulation of companies, to eliminate unnecessary bureaucracy and to ensure the greatest respect for the principle of minimum regulation. The consultation document proposes:

To strip out obsolescent and over-complex provisions and repair defective ones. ... We need clear and simplified arrangements which, starting from first principles, better capture the balance of obligations, protections and responsibilities which are required to underpin the modern marketplace so as to ensure that the participants can be confident about fair dealing.⁸⁷

The review is also considering a reformulation of the duties imposed upon directors. In the manner of a nineteenth-century code, these duties would be taken from the common law and laid out in a single statutory statement to make the principles clear and accessible.

All stages of the review will enjoy full public consultation, through representative working groups and the regular publication of key findings for comment. All proposals will undergo economic analysis to ensure that the new law strikes an appropriate balance between regulation and facilitation.

Patents – sensitivity in legislation

Another example of successful legislative intervention in the free market has been the patent system, originating in the Statute of Monopolies 1624. Patents are a clear compromise of free market principles, as they prohibit businessmen from taking the obvious and profitable step of copying other people's ideas. However, the compromise is entirely necessary. Without patents, inventors would have little incentive to innovate and technological advance would be substantially impaired.

The modern patent system is again a product of the Industrial Revolution. Before the formation of the Patent Office in 1852, the administrative burden of obtaining

⁸³ Insolvency Act 1986, s 213.

⁸⁴ *ibid* s 214.

⁸⁵ Law Commission Report No. 261, *Company Directors: Regulating Conflicts of Interests and Formulating a Statement of Duties*. This development may already have been pre-empted by the common law. See *Re D'Jan of London Ltd* [1994] 1 BCLC 561; (1994) 110 LQR 390.

⁸⁶ *Re City Equitable Fire Insurance Co* [1925] Ch 407.

⁸⁷ DTI Consultation Paper: *Modern Company Law for a Competitive Economy* (1998) para 3.8.

patents was overwhelming. Applications demanded attendance at 10 different offices, and the process had to be carried out separately in Scotland, Ireland and England and Wales.⁸⁸

The reforms of 1852 have transformed the patent system into an outstanding engine for commercial innovation. Last year alone almost 29,000 applications were made. Four features in particular have been instrumental in achieving appropriate regulation.

First, patents do not create wholly controlled monopolies. They confer on their owners the narrower benefit of exclusive commercial exploitation for a duration limited to twenty years. Even during the currency of the patent, members of the public⁸⁹ are free to conduct experiments on the patented invention.⁹⁰ The law only intervenes if the copier derives commercial advantage.

Secondly, the patent application procedure ensures that full technical details of inventions are disclosed to the public.⁹¹ Not only does this requirement ensure that patent applications are genuine, but crucially it facilitates technological advance. As Grove J said in a case of 1884:

[The applicant] is bound so to describe it in his specification as that any workman acquainted with the subject ... would know how to make it; and the reason of that is this, that if he did not do so, when the patent expired he might have some trade mystery which people would not be able actually to use in accordance with his invention (although they had a right to use it after his invention had expired), because they would not know how to make it.⁹²

Thirdly, patents protect only novel and non-obvious inventions.⁹³ Applications are subject to intensive review, by the Patent Office before grant and by judicial scrutiny after grant in any infringement proceedings.⁹⁴ This confines the rights of commercial exploitation to meritorious cases, and prevents abuse by those seeking monopoly rights on the basis of no real innovation.

Fourthly, patents have been applied in a manner that is friendly to commerce. In particular, the judge-made rules of construing patent specifications display commercial sense and subtlety. Judges do not read patent documents literally, but search for their *pith and marrow*. In the leading case,⁹⁵ Lord Reid said:

Claims are not addressed to conveyancers: they are addressed to practical men skilled in the prior art, and I do not think that they ought to be construed with that meticulousness which was once thought appropriate for conveyancing documents.

That was under previous legislation but is still the modern approach of the judges.

The patent system is strongly marked by commercial sensitivity at all levels, both in its legislative infrastructure and in its judicial application.

88 H. Dutton, *The Patent System and Inventive Activity During the Industrial Revolution* (Manchester: Manchester University Press, 1984), ch 5. Charles Dickens portrayed the arduous process and its expense in his *Tale of a Poor Man's Patent*.

89 Patents Act 1977, s 60(5)(a).

90 *ibid* s 60(5)(b).

91 This has been the case since the great reforms of 1883.

92 *Young v Rosenthal* (1884) 1 RPC 29.

93 Patents Act 1977, s 1.

94 *ibid* s 74.

95 *Rodi & Wieneberger AG v Henry Showell Ltd* [1969] RPC 367.

Equity: sensitivity in judicial rule-making

In addition to these major legislative schemes, the continuing role of the judiciary in commercial regulation must not be overlooked. There is strong international confidence in our specialised Commercial judges.⁹⁶

Of particular interest is the way in which the courts have developed principles of equity – an institution originally conceived to mitigate unfairness in the common law – in the commercial context. In the *Romalpa* case, in 1976,⁹⁷ the Court of Appeal was asked to adjudicate on the ownership of aluminium foil which had been sold to a company on the following terms:

The ownership of the material to be delivered will only be transferred to purchaser when he has met all that is owing to seller. ... Until the moment of full payment of what purchaser owes seller purchaser shall keep the objects in question for seller in his capacity of fiduciary owner.

These terms were intended to prevent ownership from passing to the buyer until the price had been paid. This turned out to be important security, as the buyer went into receivership before making payment. The seller's only effective means of recovery was to rely on these terms and claim ownership of the foil it had supplied.

The court accepted that the seller could claim the aluminium remaining in the buyer's yard. This proved a milestone in the law of secured credit, giving suppliers the opportunity to protect themselves against the insolvency of their buyers without the need for formal security in the shape of a mortgage or a charge. Since the original *Romalpa* decision in 1976, there has been a flourishing case law in this field.⁹⁸

The court also accepted that the seller was entitled to claim ownership of £35,000 in the buyer's bank account representing the proceeds of foil that had been sold on to third parties. The claim was revolutionary, as orthodoxy suggested that the seller's claim to ownership of the foil had lapsed in favour of a bare contractual claim to the proceeds of sale. Yet the court looked to principles of equity to uphold the seller's claim. It has long been established that when a party is appointed to serve the interests of a principal, whether as a trustee or an agent, he must not take unauthorised profits from the transaction. If he does, the profits belong in equity to the principal. In *Romalpa*, the court took this rule and set it in a commercial context, holding that the terms of the sale contract created a fiduciary relationship, and that any profits obtained by the buyer would belong in equity to the seller.

This is not the only example of judicial creativity in the commercial world. In company law, judges have relied on the law of fiduciaries to regulate company management, prescribing minimum standards of honesty and loyalty with which directors must comply. In these ways, the judiciary has played an important and creative role in commercial regulation even during the twentieth century.

⁹⁶ For a positive appraisal of the sensitivity of the specialised commercial judiciary in England, see R. Austin, 'Commerce and Equity – Fiduciary Duty and Constructive Trust' (1986) 6 OJLS 444. A very different system of judicial specialisation operates in North American jurisdictions, and this has been criticised: see, for example, H. Butler and L. Ribstein, 'Opting Out of Fiduciary Duties: A Response to the Anti-Contractarians' (1990) 65 Wash LR 1.

⁹⁷ *Aluminium Industrie Vaassen BV v Romalpa Aluminium Ltd* [1976] 1 WLR 676.

⁹⁸ Key cases have been *Re Bond Worth Ltd* [1980] Ch 228, *Re Peachdart Ltd* [1984] Ch 131, *Armour v Thyssen Edelstahlwerke AG* [1991] 2 AC 339, and *Compaq Computer Ltd v Abercorn Group Ltd* [1993] BCLC 602.

Conclusion – commercial law into the future

In this brief review, we have seen that the unique English system, without any distinct corpus of commercial rules, has proved an outstanding success. In particular, three features have marked the law's approach since it began the process of commercialisation in the eighteenth century: facilitation, integration and regulation. The law has *facilitated* trade by recognising the effect of commercial agreements and practices, giving effect to the intentions of contracting parties to support the free market economy. The law has *integrated* key mercantile customs into its structure, establishing coherent and predictable legal frameworks in a number of areas important to business, from bills of exchange to sales of goods. The law has also *regulated*, particularly in recent times, prescribing rules to ensure that the free market operates with the greatest internal efficiency and without detriment to the community at large.

Each of these features has assumed its greatest importance at different stages of legal development. Facilitation and integration were early objectives, as the common law courts had to reform their rules and procedures to accommodate a body of mercantile custom which had for centuries been regarded as a separate institution. Regulation is a more recent phenomenon, as increasingly sophisticated economic theories of markets have called for measures of state intervention, and governments have become more eager to act in the interests of market openness and fairness.

Yet all three processes remain relevant to the law-making process as we begin the twenty-first century.

The law continues to facilitate, by encouraging flexible frameworks that give traders the freedom to set their own standards of conduct. In the last five years, the approach to dispute resolution has been undergoing radical change to reduce the burden of litigation on business. Court procedures have been streamlined to enhance the speed and effectiveness of litigation, and in a broader context, other procedures such as strengthened arbitration,⁹⁹ and mediation,¹⁰⁰ and alternative dispute resolution, are being actively promoted.

The law continues to integrate, particularly in the international arena. Legal co-operation within the European Union has been considerable, and Community Directives have now harmonised large areas of commercial practice, from public company mergers¹⁰¹ to the use of unfair terms in consumer contracts¹⁰² and the mutual recognition of professional qualifications.¹⁰³ In the field of intellectual property, the European Patent Office in Munich has provided an invaluable service to business since its inception in 1973. It allows inventors to secure patents in every Member State with just a single application, and the Office now receives

99 As a result of the Arbitration Act 1996, the arbitration mechanism has been substantially expedited, parties have been given greater freedoms to fix their own procedures, and arbitration has been asserted as a true source of dispute resolution independent of the courts.

100 In June 1996 the Commercial Court issued a practice statement encouraging judges to adjourn proceedings at an early stage if alternative dispute resolution, particularly mediation, would be appropriate.

101 The 1st Company Law Directive of 1968 (Directive 68/151) made provision for compulsory disclosure of company information, registration, pre-incorporation contracts and ultra vires and exhaustive grounds of nullity. To date, a further four Company Law Directives have been adopted and implemented, two are awaiting implementation and five remain under negotiation.

102 Directive 93/13 (Unfair Terms in Consumer Contracts).

103 See especially Directive 89/48 on the mutual recognition of diplomas, and Directives 77/249 and 98/5 on legal qualifications.

over 100,000 applications every year. Global co-operation is also an increasing reality. I have emphasised the valuable work of the International Chamber of Commerce in formulating standard practices, and the importance of the Hague Rules in setting standards for international shipping. International instruments for the harmonisation of trade are a striking feature of the modern legal scene. In July, the International Law Association met in London to discuss proposals for international conflict of laws rules, and next month in Geneva the World Intellectual Property Organisation will consider an international instrument for the protection of audiovisual performances.

The law also continues to innovate through regulation. The limited liability partnership and the current Company Law Review are prime examples of an increasingly sophisticated law reform process led by government in consultation with the Law Commission and business.

Thus the law has been and is an essential engine for trade. By devising a principled reconciliation of laissez-faire and excessive regulation, giving scope for market freedoms while prescribing firm rules when required, English law has been the creative mediator to resolve diverse interests into a congruent whole.

In Search of the Responsible Subject: History, Philosophy and Social Sciences in Criminal Law Theory

Nicola Lacey*

This paper examines the way in which English criminal law's conception of responsibility has changed since the eighteenth century, and explores the relationship between changes in legal framework, changes in processes of criminalisation and punishment, and broader social, political and economic changes. It argues that the development of ideas of individual responsibility for crime are responses to problems of co-ordination and legitimisation faced by systems of criminal law, and that these problems can be expected to change according to the environment in which the system operates, with important factors including the distribution of political interests and economic power; the prevailing cultural and intellectual environment; the organisation and status of relevant professional groups and the vigour of alternative means of social ordering. Substantively, the paper explores the hypothesis that criminal responsibility has shifted from a conception founded in ideas of character to a capacity-based conception over the relevant period. Methodologically, the aim is to historicise the structure as well as the content of criminal law within a socio-theoretic framework, constructing a dialogue between criminal law theory of a doctrinal and philosophical temper and socio-historical studies of criminal justice.

At the start of the twenty-first century, it is little exaggeration to say that the question of responsibility – indeed the question of individual responsibility – stands as *the* question of normative criminal law theory. While, significantly, appellate cases in most legal systems of the common law world are as often concerned with the interpretation of the conduct element of criminal offences as with the responsibility condition or the defences,¹ a brief survey, not only of legal texts² and theoretical monographs³, but also of draft and model codes such as the English Law Commission's Draft Code⁴ and the American Model Penal Code testifies to the contemporary preoccupation with questions about the fault element

* Professor of Criminal Law, London School of Economics; Adjunct Professor, Research School of Social Sciences, Australian National University. I would like to acknowledge, with gratitude, the support of the Wissenschaftskolleg zu Berlin, where a fellowship during the academic year 1999–2000 provided me with the ideal environment in which to conduct the research on which this paper is based. My warm thanks also go to David Soikice for extensive discussion and encouragement, to participants at the Australian Society for Legal Philosophy conference in April 2000 and the Critical Legal Conference in Helsinki in September 2000 for stimulating feedback; and to Akcel Bilgrami, Simon Bronitt, Peter Cane, Hugh Collins, Lindsay Farmer, George Fletcher, Martin Loughlin, Tim Murphy, Alan Norrie, Mike Redmayne and Gunther Teubner for helpful comments on an earlier version.

- 1 As well as with the conditions under which collectivities such as corporations can be held responsible.
- 2 See eg Andrew Ashworth, *Principles of Criminal Law* (Oxford: Clarendon Press, 3rd ed, 1999); Alan Norrie, *Crime, Reason and History* (London: Butterworths, 1993); Glanville Williams, *Textbook of Criminal Law* (London: Sweet and Maxwell, 2nd ed, 1983).
- 3 See eg H.L.A. Hart, *Punishment and Responsibility* (Oxford: Clarendon Press, 1968); Michael S. Moore, *Placing Blame* (Oxford: Clarendon Press, 1998); Stephen Shute, John Gardner and Jeremy Horder (eds) *Action and Value in Criminal Law* (Oxford: Clarendon Press, 1993); Anthony Duff (ed), *Philosophy and the Criminal Law: Principle and Critique* (Cambridge: Cambridge University Press, 1998).
- 4 *Criminal Law: Codification of the Criminal Law* (Report No. 143, London: HMSO, 1985). A *Criminal Code for England and Wales* (Report No. 177, London: HMSO, 1989).

or mental conditions under which it is appropriate to hold an individual responsible for a crime.

In this paper, I want to raise both substantive and methodological questions about the treatment of responsibility in criminal law theory. The paper grows out of a long-term project which has two substantive purposes: first, I am studying the way in which English criminal law's conception of responsibility – its very conception of what it is to be a subject of criminal law – has changed since the late-eighteenth century; and second, I am exploring the relationship between these changes in legal framework, changes in processes of criminalisation and punishment, and broader social, political, cultural and economic changes. This substantive project is premised on the assumption that criminal law may usefully be examined not only as an institution in its own right but also as one index of broad social changes; it is therefore a project in social theory as much as a project in legal history.

Methodologically, I am also seeking to contribute to a general debate in legal theory: ie how do the different disciplinary resources of philosophy, history and the social sciences contribute to our understanding of law? The project here moves away from a conception of criminal law theory as founded primarily in analytic philosophy and the systematic analysis of legal doctrine. Rather, my aim is to historicise the structure as well as the content of criminal law within a broad socio-theoretic framework, drawing links between the conceptual structure or form of criminal law doctrine, the development of criminal procedure and penal practices,⁵ and the substantive social functions which criminal law and punishment have been expected to perform at different points in history. Hence I am attempting to construct a dialogue between criminal law theory of a doctrinal and philosophical temper and socio-historical studies of criminal justice.⁶

In doing so, I am understanding criminal responsibility as a practice of attribution which is specific to criminal law yet which is connected with prevailing intellectual ideas, including – though, obviously, not restricted to – philosophical theories about the nature of human beings. This specificity and practical orientation of responsibility in criminal law, I shall argue, entails a relationship between philosophical and legal conceptions of responsibility which is more oblique than is generally assumed in criminal law theory. Moreover, it makes it conceivable that multiple and philosophically inconsistent conceptions of responsibility may operate within legal practices of attribution without any necessary illogicality or incoherence in these distinctive practices. My intellectual starting point is, therefore, sceptical about the propriety of an *a priori* unitary approach to theorising criminal responsibility.

My initial hypothesis could be summed up in the following way: the development of ideas of individual responsibility for crime is at root a response to problems of co-ordination and legitimation faced by systems of criminal law; the content and emphasis of these problems can be expected to change according to the environment in which the system operates, with important factors including the distribution of political interests and economic power; the prevailing cultural and intellectual environment; the organisation and relative status of relevant

5 On the need to set the analysis of criminal law doctrine within the context of criminalisation and criminal justice more generally, see Nicola Lacey and Celia Wells, *Reconstructing Criminal Law* (London: Butterworths, 2nd ed, 1998) ch 1; Nicola Lacey, 'Contingency and Criminalisation' in Ian Loveland (ed) *Frontiers of Criminality* (London: Sweet and Maxwell, 1995).

6 For an elegant statement of a similar methodological approach in relation to public law, see Martin Loughlin, *Sword and Scales* (Oxford: Hart Publishing, 2000) ch 2.

professional groups and the array and vigour of alternative means of social ordering. This starting point implies that the most illuminating approach to criminal responsibility will be both a multi-disciplinary and a cross-institutional one.

In this paper, I shall introduce both substantive and methodological concerns of my overall project in the following way. First, I shall give a brief sketch of the (rather peculiar ...) field of analytic and normative criminal law theory; secondly, I shall set out what I see as the main problems arising in this field, considering in turn a number of methods for analysing and trying to overcome these problems; thirdly – and this will be the main part of the paper – I will give a more detailed account of my current research, setting out some of the hypotheses which my research so far has generated.

The nature of criminal law theory

Most intelligent people not already familiar with the field would probably assume that theorists of criminal law are engaged in debates about the substance of criminal law in particular systems or groups of systems: what kinds of things are, or should be criminalised; what interests or values criminal law seeks to express or protect. This sort of inquiry is indeed reflected in what might be called 'middle-range theory' – a genre which is increasingly vigorous and which is, happily, well represented in contemporary criminal law scholarship.⁷ Yet there persists an influential tradition in criminal law theory in Britain, Australia and North America whose exponents – they are as often philosophers as lawyers, by the way – are strikingly uninterested in these substantive questions. Indeed, they are generally quite happy to leave questions about the content and boundaries of criminal law to the attention of sociologists on the one hand and moral and political philosophers on the other. What really interests these theorists is rather the *form* of criminal law: in other words, the conceptual building blocks out of which legal doctrine constructs liability. This conceptual framework is often understood in terms of that familiar Latin maxim, '*actus non facit reum nisi mens sit rea*'. The maxim is generally agreed to be both inelegant Latin and a seriously misleading statement about the law – the Victorian jurist James Fitzjames Stephen famously described it as 'not so much a maxim as a minim' ...⁸ But, lawyers being conservative people (and not a little inclined to bemuse non-lawyers even if they have to resort to a bit of Latin to do so ...), the maxim has nonetheless survived more or less intact as a central feature of both academic and judicial thinking in the common law world.

⁷ See for example Jeremy Harder, *Provocation and Responsibility* (Oxford: Clarendon Press, 1992); Celia Wells, *Corporations and Criminal Responsibility* (Oxford: Clarendon Press, 1993); A.T.H. Smith, *Property Offences* (London: Sweet and Maxwell, 1994); A.P. Simester and A.T.H. Smith (eds.), *Harm and Culpability* (Oxford: Clarendon Press, 1996); Simon Bronitt, 'Rape and Lack of Consent' (1992) 16 *Criminal Law Journal* 289; Brent Fisse and John Braithwaite, 'The Allocation of Responsibility for Corporate Crime: Individualism, Collectivism and Accountability' (1988) 11 *Sydney Law Review* 468; Ngaire Naffine, 'The body bag', Ian Leader Elliot, 'Passion and insurrection in the law of sexual provocation', and Rosemary Hunter and Kathy Mack, 'Exclusion and silence: procedure and evidence', all in Ngaire Naffine and Rosemary Owens (eds) *Sexing the Subject of Law* (Sydney: Law Book Company, 1997) at 79, 149 and 171 respectively; and P. Ruah, S. McVeigh and A. Young (eds) *Criminal Legal Doctrines* (Aldershot: Dartmouth, 1997); Lois Bibbings and Donald Nicholson (eds) *Feminist Perspectives on Criminal Law* (London: Cavendish, 2000).

⁸ James Fitzjames Stephen, *A History of the Criminal Law of England* (London: Macmillan, 1883) vol II, 94.

A more accurate statement would be that the general conditions for criminal liability include both a conduct element – eg causing death, in the case of homicide – and a mental or responsibility element – eg intending to kill. So even if a statutory provision does not include a *mens rea* word such as ‘intent’, judges – so the story goes – will generally read such a word into the provision. Moreover, even if the conduct and responsibility elements are satisfied, it may be that a generally applicable defence removes liability – in cases, for example, where the defendant was defending herself against attack, or was under duress, or was suffering from a serious form of mental incapacity. This three-tiered structure is, of course, the basis of what is known as the *general part of criminal law*.

At one level, this is simply an analytic framework; it purports to explicate the general structure of criminal liability. But this is not the only reason why it interests criminal law theorists. For implicit in most discussions of the Latin maxim is a strong normative thesis about the legitimacy of criminal law. Like the presumption of innocence and principle of legality, the idea that criminal liability must involve both conduct and responsibility or fault lies at the heart of contemporary criminal law’s idea of itself as not just a system of state power and force, but a legitimate – even in some senses a moral – system. Most obviously, in contemporary discussions, the maxim connects with a cluster of values associated with political liberalism. For example, the principle that criminal liability generally requires a positive act rather than merely an omission – in contrast to some systems of the continent of Europe⁹ – appears to be premised on a libertarian idea that it is less intrusive for the state to prohibit conduct than positively to require it. Similarly, the idea that one can only be liable for an action for which one is in some sense responsible connects with fundamental ideas of fairness and of respect for individual liberty.

In fact, as we will see, members of this school of criminal law theory today tend to be primarily interested not so much in the conduct as in the responsibility condition,¹⁰ – what I shall call from now on ‘the principle of responsibility’. So how is this fundamental idea of responsibility conceived? The dominant argument is that individuals’ criminal responsibility is at root based upon their capacities and opportunities: capacities of cognition or understanding: it is only fair to punish someone who has the capacity to understand what they are doing; and capacities of volition or will: it is only justifiable to punish someone who has the capacity or, in some versions, the fair opportunity to act otherwise than they did.¹¹ In philosophical terms this capacity theory of responsibility assumes either individual free will or some version of ‘soft determinism’ – the idea that, in Strawson’s terms, our reactive attitudes such as praising and blaming could survive the truth of determinism.¹²

9 See Alberto Cadoppi, ‘Failure to Rescue and the Continental Criminal Law’ in M. Menlowe and A. McCall Smith (eds), *The Duty to Rescue* (Aldershot: Dartmouth, 1993).

10 For exceptions to this feature of contemporary theory, see Michael S. Moore, *Act and Crime* (Oxford: Clarendon Press, 1993) Paul H. Robinson, *Structure and Function in Criminal Law* (Oxford: Clarendon Press, 1997); and Douglas Husak, ‘Does Criminal Liability Require an Act’ in Duff, n 3 above, 60. These count, however, as only partial exceptions; Robinson in the sense that his account of conduct aspects sits alongside a substantial treatment of responsibility; Moore because his account of action is ultimately located within a theory of the conditions of moral responsibility for crime; and Husak for the same reason as Moore and because he rejects the idea that criminal liability requires an act.

11 See Hart, n 3 above.

12 ‘Freedom and Resentment’ (1962) xlviii *Proceedings of the British Academy* 1–25.

In legal terms, the capacity conception is generally interpreted as entailing that the paradigm case of 'mens rea' or legal responsibility is a subjective mental state.¹³ Obvious examples would be intention, foresight of harm, or knowledge of circumstances; a person accused of murder must intend to kill or cause serious harm; a person accused of rape must know that the person he rapes does not consent; a person accused of assault must foresee that her action is likely to cause some physical harm. But what about 'objective' versions of fault – eg negligence in the sense of failure to reach a reasonable standard of care or behaviour? These are reluctantly admitted to have some relevance, but are represented as empirically marginal and normatively irregular. For while, on the capacity theory, the imposition of objective standards on those with the general capacity to reach a reasonable standard can indeed be justified, the imposition of the sort of fully objective standard not uncommon in English criminal law¹⁴ is unfair because it fails to investigate whether or not the defendant indeed had the capacity and hence a genuine opportunity to reach the standard imposed.¹⁵ As for so-called 'strict liability' – liability without fault – it is generally mentioned in hushed tones as an embarrassing and uncivilised exception to the general principles of criminal law.¹⁶

The history of this predominantly subjective, capacity-based conception of responsibility, and the reasons why it has assumed such pre-eminent status in criminal law theory, are the main questions of my research. But my current historical project is premised on questions thrown up by analytic and socio-legal research on criminal law conducted over the last decade. So I want briefly to refer to that work both to set the scene for the historical project and to illustrate the methodological question which is equally part of my concern.

Questioning the unitary subjectivist theory of responsibility

The 'internal critique'

The starting point for much recent research has been a degree of scepticism about whether criminal law in fact reflects any consistent and principled idea of individual responsibility.¹⁷ In particular, scholars have focused on the intriguing way in which both judges and commentators on criminal law almost universally acknowledge the inaccuracies and even obscurity of the Latin maxim yet almost simultaneously both appeal to it and assume its importance. This disjuncture has been extensively analysed in what we might call an 'internal critique' of criminal

13 See for example Williams, n 2 above. The subjective approach was recently endorsed by the House of Lords: see *B v DPP* [2000] 1 All ER 833.

14 See, for example, *Elliot v C (A Minor)* [1983] 2 All ER 1005; the imposition of objective tests irrespective of capacity also characterises several defences: on duress, for example, see *R v Bowen* [1996] 4 All ER 837.

15 Hart, n 3 above, chapter 4, argues that negligence liability is consistent with capacity responsibility because the defendant is held responsible for their failure to attain a (reasonable) standard. This argument only goes through, however, if we assume that the defendant indeed had the capacity to attain the relevant standard. It therefore justifies only a 'partially subjective' negligence standard, and not what I shall call the 'strongly objective' standard pertaining in much contemporary English criminal law.

16 This is notwithstanding powerful arguments to the effect that responsibility for outcomes may be morally justified irrespective of direct capacity-responsibility: see eg Tony Honore, 'Responsibility and Luck' in his *Responsibility and Fault* (Oxford: Hart Publishing, 1999).

17 For discussion of this tendency in recent criminal law theory, see Anthony Duff's Introduction to Duff, n 3 above, 1.

law's realisation of the maxim.¹⁸ For example, take the presumption that all or at least the vast majority of serious offences require proof of full subjective responsibility. If we take even what is surely a paradigm case – the crime of murder – we find that the requirement of responsibility, though undoubtedly subjective, falls somewhat short: in English law, a person can be liable for murder even if they intend not death but a lesser degree of harm. Indeed, until as recently as 1957, someone who unintentionally killed while in the course of committing another serious offence could automatically be convicted of murder.¹⁹ More obviously, in the frequent case of offences where the conduct element is a complex one, the responsibility requirement often applies only to some elements of the conduct. So, to take an example, one can be liable for the offence of assault occasioning actual bodily harm even if one did not intend or foresee any harm.²⁰

Furthermore, English criminal law today – like most other systems in the common law world – has many offences which can be committed with strongly objective negligence or indeed without fault. The offence of manslaughter, which carries a maximum penalty of life imprisonment, can be committed negligently; and a vast array of offences – not only in the so-called 'regulatory' areas of licensing, industrial safety or environmental health but also, for example, in the area of public order – require no proof of '*mens rea*' at all.²¹ Two scholars recently studied all indictable offences on the statute book in England and Wales and found that no fewer than 50 per cent had an incomplete responsibility requirement, while 40 per cent made some modification of the presumption of innocence.²² In this context, the internal critique has been geared to suggesting – to put it very crudely – that the principle of subjective responsibility might reasonably be seen as not so much a logical but rather an ideological feature of contemporary criminal law.

This internal critique which I have described finds its roots in methods of analysis which are typically – though not in my view distinctively – legal. Certainly, it confines its data to the legal system. Although I have certainly been a participant in its development, I have gradually come to the conclusion that the basis of the method is unduly limited, and that one needs to step outside the boundaries of legal doctrine to throw further light on the significance of the fact that the responsibility principle is so extravagantly honoured while being so regularly breached. So in more recent years I have been drawing on the social sciences – mainly on the sociology of criminal justice – to further the analysis.²³

Sociological perspectives

There are various ways in which one can draw on sociological evidence to shed further light on the significance of the responsibility principle in criminal law. An obvious example would be data on patterns of criminal enforcement, which show that crimes with no responsibility requirement or a partial responsibility

18 See for example Mark Kelman, 'Interpretive Construction in the Substantive Criminal Law' 33 *Stanford Law Review* (1981) 591; Nicola Lacey, 'A Clear Concept of Intention?' (1993) 56 *Modern Law Review* 621.

19 A legal position which still pertains in the USA: see George Fletcher, *Rethinking Criminal Law* (Boston and Toronto: Little, Brown, 1978) 307–319.

20 *R v Parmenter* [1991] 4 All ER 698.

21 See Lacey and Wells, n 5 above, ch 2.III.a and ch 5.II.b.

22 Andrew Ashworth and Meredith Blake, 'The Presumption of Innocence in English Criminal Law' (1996) *Criminal Law Review* 306.

23 See for example Nicola Lacey, 'Contingency and Criminalisation' in Loveland, n 5 above.

requirement are empirically dominant – ie they are over-represented in enforcement practice relative to their numbers on the statute book.²⁴ Less obviously, research shows that the distinction between crimes requiring proof of fault and those of strict liability, which is regarded as foundational in legal doctrine, is much less clear in practice. Here, pressures on defendants to plead guilty, or the limited scope of criminal defences, often mean that those convicted of *mens rea* crimes are less than fully responsible in the relevant sense. Conversely, enforcement patterns for strict liability crimes suggest that scarce prosecution resources tend to be rationed along fault-based lines: only those seen by enforcement officials as truly responsible are likely to be prosecuted.²⁵ Another rich sociological seam is the investigation of assumptions and prejudices which shape the interpretation of the responsibility requirement for different classes of defendant. For example, there is convincing evidence that judges are more likely to accept arguments such as diminished responsibility, which question full mental responsibility, in relation to female defendants rather than in relation to male defendants – in other words, to interpret female offending as ‘mad’ rather than ‘bad’.²⁶

Historical approaches

More recently, however, my interests have turned to history, which seems to offer a further purchase on the contemporary significance of the principle of responsibility. What pushed me in a historical direction was my increasing bemusement about a certain style of analysis which is very common in contemporary Anglo-American criminal law theory. This style – which is strongly associated with analytic philosophy – could be caricatured in the following way:

The conceptual structure of criminal responsibility consists in volitional and cognitive capacities (described above ...); any departure from this conceptual structure entails logical or moral confusion: it mistakes the concept of responsibility for something else, and in doing so it makes a moral mistake in purporting to justify the attribution of responsibility to those who are not ‘truly’ responsible.²⁷

This sounds like a metaphysical claim: a claim which does not entertain the possibility that particular conceptions of responsibility might have distinctive conditions of existence – intellectual, cultural, economic, social, political or otherwise. One way of scrutinising such an apparently transcendent claim about the structure of criminal responsibility is to historicise it: another, of course, would be to study it comparatively. A historical approach is central to Alan Norrie’s important *Crime, Reason and History*,²⁸ though the primary focus of Norrie’s book is a critical analysis of the general part of contemporary English criminal law. And both historical and comparative methods are in evidence in a book which, though published over twenty years ago, still stands as one of the most intellectually

²⁴ See n 22 above.

²⁵ W.G. Carron, ‘White-collar Crime and the Enforcement of Factory Legislation’ (1970) 10 *British Journal of Criminology* 383; K. Hawkins, *Environment and Enforcement* (Oxford: Clarendon Press, 1983); Lacey and Wells, n 5 above, chs 1 and 5.

²⁶ See for example Hilary Allen, *Justice Unbalanced* (Milton Keynes: Open University Press, 1987).

²⁷ Note the slippage between analytic and normative voices – an interesting topic in itself, but not one which I can pursue here. A powerful example of this approach is Moore, n 3 above.

²⁸ n 2 above; see also Lindsay Farmer’s *Criminal Law, Tradition and Legal Order* (Cambridge: New York: Cambridge University Press, 1996) chs 1 and 5; S. McVeigh and P. Rush, ‘Cutting our Losses: Criminal Legal Doctrine’ in Rush, McVeigh and Young, n 7 above.

challenging treatises on criminal law – George Fletcher’s monumental *Rethinking Criminal Law*.²⁹ Fletcher’s and Norrie’s are, however, rare voices; moreover, Fletcher’s approach to the development of conceptions of responsibility is a steadfastly legal one, which shows relatively little interest in the interaction between the broad legal developments which he traces and their political, cultural and socio-economic context.

The lack of any historical dimension to most contemporary Anglo-American criminal law theory³⁰ is striking, but it is not complete. There is a weak historical thesis which is quite regularly acknowledged – at least with a nod – by criminal law theorists. It could be called ‘The Story of Enlightenment’ or perhaps ‘of Modernity’.³¹ This story acknowledges that contemporary conceptions of individual responsibility are strongly associated with a set of ideas about human being and a set of values which finds expression, variously, in the work of philosophers such as John Locke, Immanuel Kant and Jean-Jacques Rousseau and which emerged in European public consciousness in the eighteenth century: the responsible subject of criminal law today is a certain kind of citizen, and a rational, self-determining moral agent.

Obviously, this story has some plausibility. Equally obviously, it is incredibly crude. In particular, it fails to account for three things. First, it does not account for the fact that other ideas of responsibility – notably the idea of responsibility as founded not in an individual’s volitional and cognitive capacities but in her character or in the disposition manifested in her conduct, to be found in philosophies as different as those of Aristotle, Spinoza and Hume – continue to have some resonance in modern Anglo-American law.³² Second, thus, it doesn’t account for the fact that the subjective, capacity conception is only one among several possible interpretations of human responsibility in post-Enlightenment traditions.³³ Thirdly, it fails to account for the fact that, as I shall try to show, the current idea of subjective, capacity responsibility comes to prominence in English legal thinking only in the early nineteenth century, and in English law a great deal later than that. It is therefore not, as a casual reader of contemporary criminal law theory might be forgiven for thinking, a long-standing feature of common law thought.

29 n 19 above; cf Oliver Wendell Holmes, *The Common Law* (Cambridge, Mass.: Belknap Press, first published 1881, Mark DeWolfe Howe ed, 1963) ch II.

30 Interestingly, penal theory is somewhat different – a point to which I shall return.

31 The relationship between conceptions of responsibility and modernity is explored in more detail in Nicole Lacey, ‘Responsibility and Modernity in Criminal Law’ (2001) 9 *Journal of Political Philosophy* 1.

32 See Michael Bayles, ‘Character, Purpose and Criminal Responsibility’ (1982) 1 *Law and Philosophy* 5 (drawing on Hume’s philosophy); John Gardner, ‘The Gist of Excuses’ 1 *Buffalo Criminal Law Review* (1998) 575; Dan M. Kahan and Martha C. Nussbaum, ‘Two Conceptions of Emotion in Criminal Law’ (1996) 96 *Columbia Law Review* 269; Nicola Lacey, ‘Partial Defences to Homicide’ in Andrew Ashworth and Barry Mitchell (eds), *Rethinking English Homicide Law* (Oxford: Clarendon Press, 2000). On character and responsibility more generally, see Ferdinand Schoeman (ed), *Responsibility, Character and the Emotions* (Cambridge: Cambridge University Press, 1987); Moira Gatens and Genevieve Lloyd, *Collective Imaginings* (London: Routledge, 1999) chs 3 and 6 (on Spinoza and responsibility). Echoes of the character conception appear especially strong in objective versions of criminal responsibility where the relationship between the (inaptly termed) ‘mens rea’ element and the specific ‘actus reus’ is attenuated, suggesting that responsibility resides in the negligent attitude itself. For further discussion, see Lacey, n 31 above.

33 For discussion of the multiplicity of traditions encapsulated in the Enlightenment(s), see Fania Oz-Salzberger, ‘New Perspectives on the History of the Enlightenment: Can Disparate Perspectives Make a General Picture?’ *Tel Aviver Jahrbuch für deutsche Geschichte* (Fall 2000).

As well as its apparently metaphysical pretension, a further feature of the influential analytic approach presents an interesting object of historical inquiry. This is a related set of assumptions about the proper nature of both criminal law and of attempts to rationalise or theorise it as a body of doctrine. A suitable caricature would be this:

The task of criminal law theory is to work out a refined, general, systematic framework for criminal law; this can only be done by means of the elaboration of general principles of liability structured through concepts such as responsibility – concepts which must be used in the same way throughout the law.³⁴

In this instance, the theorist's methodological claim is quite often decorated with vague historical references to 'early periods' in which the organisation of criminal law was 'arbitrary' or 'chaotic'; an era in which criminal law's unfortunate subjects did not benefit from the enlightened achievements of certainty, predictability or coherence guaranteed by a system based on a principled 'general part'. In a curious mix of analytic philosophy and a Whiggish (or perhaps historically 'ethnocentric') theory of history, the development of criminal law is evoked as a long struggle of heroic (preferably philosophically literate) intellectuals to convince recalcitrant and conservative judges and legislators of the merits of a principled system. This struggle, so the story goes, has long been won at the level of treatises on criminal law but not – though this is too embarrassing to be consistently admitted – at the level of criminal law itself. At a more concrete level, these themes are reflected in the arguments of proponents of the codification of English criminal law.

Quite apart from my doubts about this linear, teleological view of history, I am interested in a number of things about this story. First, it makes a philosophical assumption that only arguments which can be subsumed within a general schema of principles are good, rational arguments.³⁵ Second, it therefore assumes that only a rich set of general principles can guarantee a rational and systematic criminal law. Third, it implies that the focus of criminal law theory should be mainly on the form of liability – the 'general part' – and not on the substance of offences – the so-called 'special part' of criminal law.³⁶ Each of these things is, in my view, open to question, particularly in view of the expanding terrain occupied by, and increasingly diverse social roles played by, criminal law from the mid-nineteenth century onwards.

With these problems in mind, a couple of years ago I undertook what might be called a pilot study to begin to trace the history of the 'general part' of criminal law.³⁷ I selected three influential and substantially comparable texts – all of them legal commentaries, but all written at least in part with an audience of lawyers and legislators in mind, and – obviously – drawn from different periods. I then

34 This claim that criminal law must be based on a 'general part' is an idea which, incidentally, gripped the legal imagination on the continent of Europe earlier than was the case in the common law world, and it is central to most modern projects of codification. Several of the assumptions contained in this caricature can be found in K.J.M. Smith's *Lawyers, Legislators and Theorists* (Oxford: Clarendon Press, 1998) discussed in detail below: they characterise most contemporary English textbooks of criminal law as well as theoretical treatises such as that of Moore, n 3 above. On the 'quest for a general part', see also Fletcher, n 19 above, ch 6.

35 For a critique of this view, see John Gardner, 'On the General Part of Criminal Law', in Duff, n 3 above, 205.

36 This focus on the form of criminal law and on its general part characterises even Norrie's critical text: n 2 above.

37 See Nicola Lacey, 'Contingency, Coherence and Conceptualism: Reflections on the Encounter between "Critique" and "the Philosophy of the Criminal Law"' in Duff, n 3 above, 9.

compared their treatment of what we would today call 'the general part'. The texts were William Blackstone's *Commentaries on the Laws of England*, published in the 1760s;³⁸ James Fitzjames Stephen's *History of Criminal Law* and his *Digest of Criminal Law* (a Draft Criminal Code) from the 1880s;³⁹ and Glanville Williams' hugely influential *Criminal Law: The General Part*, first published in 1953.⁴⁰

A structural as well as a content analysis of these three texts produced some (gratifyingly!) striking contrasts. Most obviously, the idea of a general part is missing from Blackstone's text and – more significantly – exists only in sketchy form in Stephen's; while each of them spends only about 10 per cent of their text considering rules and principles applicable across offences (including not only the requirements of conduct and responsibility but also defences and matters such as the nature of penalties, inchoate offences such as attempt and liability for secondary participation), Glanville Williams' text – which provides the model for almost all subsequent commentaries and many attempts at codification – almost exactly reverses the proportion. For both Blackstone and Stephen, it is the so-called 'special' part – the criminal offences themselves – which are, as it were, where the theoretical action is. Their observations about general conditions of liability are focused on the nature of voluntary conduct and defences rather than on responsibility; these observations are very brief, and are elaborated mainly in relation to particular offences. Even Stephen's earlier *A General View of the Criminal Law of England*,⁴¹ contrary to what today's reader might assume from its title, has only one chapter, spanning a mere 24 of the book's nearly 500 pages, which discusses individual responsibility and which corresponds with the 'general part' in Williams' sense.⁴²

This, I would argue, does not entail that either Blackstone's or Stephen's account lacks rationality, coherence or system. Indeed, if we look at the ways in which they organise their material, we perceive a perfectly systematic approach, as well as learning a great deal about how they saw their field. Let me take just one example. For Blackstone – a great champion of the common law, and writing as statute law was set to become an increasingly significant force – the challenge was to present the common law of crime as rational and reasonably systematic in terms not of the arcane procedural distinctions for which common law was (in)famous but rather of the interests it protected and the values it upheld. Blackstone therefore presented criminal law in a way which spoke to these issues, grouping offences not on terms of their distinctive procedural characteristics but in terms of substantive concerns which could be seen as adding up to a coherent sense of 'public wrong': offences against God and religion, offences against the king, offences against public order, offences against the person and property. He then elaborated these in relation to exemplary fact situations which spoke to the important legal issues of his time.

From the point of view of late-twentieth-century legal philosophy, Blackstone's *Commentaries* certainly do look a bit chaotic; there are few articulations of

38 1765–69 (Chicago: University of Chicago Press, 1979).

39 Stephen, n 8 above.

40 (London: Stevens and Sons, 2nd ed, 1961). I am grateful to Tony Smith for alerting me to the fact that Williams originally planned a four-volume work, with additional volumes on offences against property, persons and the state. The success of the one volume as a self-standing work is, however, highly significant.

41 (London: Macmillan, 1863).

42 *ibid* ch III; the book as a whole, which is a precursor to the *History*, deals mainly with the historical development of criminal law, with the definition and classification of crimes, with criminal procedure and punishment, and with the development of criminal legislation; like the *History*, it concludes with a description and discussion of some exemplary trials.

anything we would recognise as general principles, let alone any sustained reflection on the nature of voluntary conduct, individual responsibility or the specific excusing conditions which might undermine ascriptions of responsibility. But Blackstone's exemplary method of elaboration and analysis makes a great deal of sense in its own terms. (It is also, curiously, strongly resonant with Wittgenstein's description of language games;⁴³ in a sense, Blackstone tries to introduce us to the common law by bringing it alive as a distinctive practice or way of life). He is – from a late-twentieth-century perspective – spectacularly uninterested in the question, 'what makes it fair to hold an individual responsible for a crime'. But, as I shall argue below, this should not come as any surprise when we reflect upon criminal justice institutions and socio-political conditions in mid-eighteenth-century England.

I could multiply these contrasts between the eighteenth, late-nineteenth and mid-twentieth-century texts and have written about them in more detail elsewhere.⁴⁴ But for the purposes of this paper, suffice it to say that the texts provide powerful evidence of a broad historical trajectory of conceptions of criminal responsibility in English law. To put it simply, the problem of individual responsibility was not a big issue for Blackstone, whose analysis might be seen as showing a deep confidence in the power of judge and jury to recognise criminal conduct when they saw it. Yet more significantly, this also seems broadly to be true for Stephen. Certainly, there emerge within his framework two notable ideas absent from Blackstone's: first, a vision of the human subject as a rational actor capable, therefore, of being deterred by a systematic and well publicised system of criminal law proscriptions and sanctions; second, in his discussion of insanity (which is very extensive as compared with his treatment of other issues of responsibility⁴⁵), a concern with the lessons which medicine and psychology imply about the limits of human responsibility. But only by the 1950s has the question of individual, let alone subjective, responsibility *as a question of agency and of individual fairness* become the primary focus for criminal law commentaries.

Clearly, we have to be cautious about how much we read from the structure of criminal law commentaries – even those written explicitly for practitioners – about the practice of criminal law itself. Such commentaries reflect, nonetheless, something important: they show us how influential exponents (some of them also practitioners) believed they could best make sense of criminal law for a legal as well as a broader audience. As a first, highly speculative hypothesis, therefore, I would argue that the specific shape of their changing views of how to organise their material – of what was most significant, and why – might suggest that the nature of the legitimacy and co-ordination problems faced by English criminal law has changed hugely during the period. It would be stretching matters to suggest that the eighteenth-century criminal law lacked a conception of responsibility. Clearly, Blackstone's commentaries are already grappling with ideas of will which form the bedrock for the notion of capacity responsibility in late-modern criminal law. But the idea that the criminal trial of that time was an investigation of individual responsibility in anything like the sense that it is today would be preposterous. In fact, evidence about the trial process at that time gives vivid

43 Ludwig Wittgenstein, *Philosophical Investigations* (Oxford: Basil Blackwell, tr G.E. Anscombe, 1963).

44 See n 37 above.

45 Stephen's discussion of insanity occupies 62 pages of the second volume of the *History*, n 8 above – more than twice as much as his general discussion of responsibility.

support to the argument that criminal law was still operating to a significant extent with something akin to Fletcher's idea of 'manifest criminality'.⁴⁶

This claim needs, of course, to be made with care. The self-informing jury as witnesses, direct or indirect, of danger immediately visible to the community, and along with it the early form of manifest criminality typified by catching someone in the act of stealing, were already, in the eighteenth century, a thing of the past. But it had decisively *not* been superseded by a conception of subjective criminality, and the role of local knowledge of a more indirect kind was still central to the process of criminal judgment. This raises the possibility of a model of criminal responsibility intermediate, both historically and conceptually, between manifest and subjective criminality: between an exculpatory system which assumed the defendant's responsibility and a fully inculpatory system in which that responsibility is a matter of proof at trial. On this model, the jury's increasingly indirect knowledge of the criminal conduct was mediated by assumptions and judgments about the defendant's character – assumptions which were crucial to the conduct of what was still an essentially exculpatory process.⁴⁷ Criminality was no longer directly perceived; but shared assumptions about good and bad character were increasingly relied upon as evidence capable of making the defendant's guilt manifest, reinforcing the community consensus about many norms of criminal conduct which continued to be shared in the late eighteenth century. This model, operating in a world in which professional policing and prosecution and an elaborate law of evidence had yet to be developed,⁴⁸ worked on the basis of lay evaluation of normative, character-based – rather than subjective or psychological – evidence and assumptions about the individual defendant. The character conception did not make the sort of '*actus reus/mens rea*' distinction common to modern doctrine. It had, however, more than a passing resemblance to normative philosophical ideas of character responsibility. Its distinctive qualities are suggested by the frequent use of evaluatively laden 'responsibility' terms – the most obvious would be 'maliciously' – in elaboration of common law offences.⁴⁹

In the context of the wide array of capital penalties which were relatively rarely executed, it has been persuasively argued that the eighteenth-century criminal trial was essentially a process for selecting those to whom the draconian penalties would be applied.⁵⁰ Every aspect of the trial process was based on the assumption that the jury could recognise criminality when they saw it: verdicts of not guilty were rare;⁵¹ there was nothing approaching the modern presumption of innocence; and most of the argument concerned factors which should mitigate the full rigour of the penal law. Hence the local knowledge of experienced and articulate juries, afforded by evidence which was predominantly concerned with the defendant's

46 Fletcher, n 19 above, ch 3. The evidence about the policing, prosecution and trial process which underpins this thesis is discussed in more detail in n 31 above.

47 I am particularly grateful to Lindsay Farmer for suggestions and discussion on this point; see also Farmer, n 28 above, 119 ff.

48 On the development of policing, see Leon Radzinowicz, *A History of English Criminal Law and its Administration from 1750: vol. IV: Grappling for Control* (London: Stevens and Sons, 1968) chs 1–7; and on evidence, John H. Langbein, 'Historical Foundations of the Law of Evidence' *Columbia Law Review* (1996) 1168.

49 Jeremy Horder has recently argued that the 'malice principle' constitutes a discrete and under-recognised aspect of '*mens rea*' in English criminal law: 'Two Histories and Four Hidden Principles of *Mens Rea*' (1997) 113 *Law Quarterly Review* 95.

50 See J.M. Beattie, *Crime and the Courts in England 1660–1800* (Princeton: Princeton University Press, 1986) 230 ff.

51 See Peter King, 'Punishing Assault: The Transformation of Attitudes in the English Courts' (1996) XXVII *Journal of Interdisciplinary History* 43, 50–51.

reputation and character in the local community, and the defendant's own statement to the jury at the conclusion of the trial, formed the centrepiece of the proceedings. Responsibility in our sense today was either assumed or irrelevant: the key question was whether the guilty defendant was a corrigible member of the law-abiding community. This turned on the character and disposition which the alleged offence, along with other pieces of local information and the testimonies of character witnesses, suggested.⁵²

History in the service of legal and social theory

The features of the eighteenth-century system which I have mentioned, along with the evidence of its gradual transformation provided by my pilot study, suggest three areas of particular interest. First, the changing relations between individual citizens and the nation state and its institutions – indeed the emergence of a conception of increasingly egalitarian citizenship throughout the nineteenth century – must surely have affected the salience of individual responsibility as a focus of criminal trials. Second – particularly given Stephen's significant emphasis on mental incapacities – the growing influence of medical sciences and psychology on criminal law's understanding of human behaviour and the gradual emergence of the idea that the defendant's interior world might be an object of medical and legal knowledge seems likely to have had a decisive impact, with the development of the insanity defence gradually influencing broader ideas of responsibility.⁵³ Thirdly, we might expect to see some interaction between the rapidly expanding scope of criminal liability during the nineteenth century – an enlarged conception of the social functions of criminal law concomitant with the enlarged functions and capacities of the administrative state,⁵⁴ the accompanying rationalisation and formalisation of the prosecution, trial and penal processes – and the doctrinal arrangements for ascribing responsibility to those accused of crime. For as the terrain of criminal law expanded and its social functions diversified, it would seem likely that its normative foundations, which had been adequate in underpinning the practice of judgment over a smaller range of issues, would come under increasing strain. As Alan Norrie has argued, this problem of legitimation would have been of particular importance in the context of perceived threats to the stability of the property relations which lay at the heart of the emerging capitalist economy.⁵⁵

52 Another significant feature of the eighteenth-century system has to do with the apparently low intensity of criminal prosecution relative to offending behaviour. This seems likely to have been born not only, negatively, of the lack of professional police and a widespread reluctance to pursue prosecution, the trouble and cost of which was largely borne by the victim, but also, positively, of a confidence in the power of local methods of informal social control, drawing on strong norms of respectability and dense networks of local knowledge of reputation, to resolve disputes and assert norms of orderly behaviour; see Beattie, n 50 above; V.A.C. Gatrell, Bruce Lenman and Geoffrey Parker (eds), *Crime and the Law* (London: Europa, 1980), essays by Lenman and Parker ('The State, the Community and Criminal Law in Early Modern Europe' at 11), David Philips ('A New Engine of Power and Authority' at 156), and Gatrell ('The Decline of Theft and Violence in Victorian and Edwardian England' at 239); King, n 51 above, 50.

53 On this question, see Roger Smith, *Trial by Medicine* (Edinburgh: Edinburgh University Press, 1981); Joel Peter Eigan, 'Lesion of the Will: Medical Resolve and Criminal Responsibility in Victorian Insanity Trials' (1999) 33 *Law and Society Review* 425.

54 See Oliver MacDonagh, 'The Nineteenth Century Revolution in Government: A Reappraisal' (1958) 1 *Historical Journal* 52.

55 Norrie, n 2 above, chs 2 and 3.

However, my initial study of course suffered two crucial deficiencies. First, from a historical point of view, it was far too crude: the gap in time between the texts was too long, and further examples would have to be examined so as to refine the analysis. Secondly, it raised far more questions than it answered. Notably, assuming that the historical development I had identified turned out on closer examination to be broadly confirmed, what was driving it? Or – to put it in less ambitious terms – what did it signify? To get at these sorts of questions, it would be necessary to diversify the texts studied, adding to the commentaries primary legal sources, para-legal sources such as model codes, and to examine a much wider variety of materials capable of shedding light on the intellectual cultural, political and economic context in which these legal ideas and debates were being formed.

In pursuing this potentially unmanageable project, I have had two very important pieces of luck. The first is that over the last decade there has been a flourishing of literature on developments in criminal procedure and punishment over the relevant period which provide a rich source of evidence for a cross-institutional analysis.⁵⁶ The second is that there are already two studies which track very closely the Anglo-American legal developments in conceptions of responsibility over the relevant period, but which do so in what one could call ‘mono-disciplinary’ – legalistic – terms, hence providing a vivid contrast to the multidisciplinary approach which I want to develop.⁵⁷ There remains, therefore, a need to bring into dialogue an ahistorical field of criminal law theory; a field of legal history which proceeds in isolation from social and criminal justice history; and a social history of criminal justice which pays scant attention to the specificities of legal reasoning, legal doctrine or even the content of criminal legal rules. Further than this, it seems worthwhile to articulate these conceptual and historical insights with some broad socio-theoretic understanding of what role criminal justice plays in the development of modern societies and of what, conversely, we can learn about the nature of our social orders from an examination of their processes of attributing responsibility and distributing punishment.

Two histories of criminal responsibility

To give an idea of how my research so far has suggested some more specific hypotheses about the interaction of legal conceptions of responsibility with broader criminal justice and social developments, let me give a brief sketch of two very different approaches. The first, which I shall call The Cultural Historian’s Tale, can be represented by a fascinating book by Martin Wiener – *Reconstructing the Criminal*.⁵⁸ Wiener analyses a wide variety of sources, ranging from novels through newspapers, parliamentary proceedings, legal and criminological texts, official reports such as those of the Criminal Law Commissions set up during the course of the nineteenth century and – though less extensively – legislation and

⁵⁶ See for example David Garland, *Punishment and Welfare* (Aldershot: Gower, 1983); David Garland *Punishment and Modern Society* (Oxford: Oxford University Press, 1990); n 50 above; David J.A. Cairns, *Advocacy and the Making of the Adversarial Criminal Trial* (Oxford: Clarendon Press, 1998); Carolyn A. Conley, *The Unwritten Law: Criminal Justice in Victorian Kent* (Oxford: Oxford University Press, 1991); Gatrell et al, n 52 above; J.H. Langbein, ‘The Criminal Trial before the Lawyers’ (1978) *University of Chicago Law Review* 263; Farmer, n 28 above.

⁵⁷ See n 19 above; Smith, n 34 above.

⁵⁸ (Cambridge: Cambridge University Press, 1991.)

case law. He argues that the idea of responsibility for crime was undergoing a radical change through the century. Early on, there was both widespread official anxiety about crime and a decisive reaction against the eighteenth-century system of threatening very severe penalties such as capital punishment which were only irregularly enforced, and in which judiciary, jury and Crown had a wide degree of discretionary power. In contrast, there emerged both a democratic concern for uniformity in the administration of criminal law and a powerful discourse of individual responsibility based on defective character.⁵⁹

In this vision, capacity was certainly relevant, but not in the way it is today. For it was not the capacity for understanding or opportunity for direct control of the criminal act itself but rather the capacity to work on one's character which was the important thing: defective, criminal character was understood as the failure to exercise general self-government or self-discipline. Furthermore, neither capacity in the sense of direct control for criminal conduct nor capacity to work on character were regarded, at least in the early part of the century, as objects of investigation at the criminal trial. Capacity was, in short, an operative assumption of the process rather than an object of legal inquiry.⁶⁰ Criminal behaviour was seen as proceeding from uncivilised, savage human nature; but through the announcement of a clear set of norms and threats, and through the intervention of the modern prison, proper habits of self-governance could be instilled into a deviant but potentially malleable population. Wiener argues that this was not necessarily a belief in 'actual' responsibility in the sense of free will; rather, it was a governmental belief that the best way to get people to conform was to treat them *as if* they were fully responsible in the sense of having the capacity to work on their characters. A plausible interpretation of the historical data is that developments in penalty – the erosion of capital punishment and transportation, the birth of the prison, an optimism about shaping human conduct via penal discipline, and, last but not least, a concern with the increasing economic costs of punishing – would have been important factors in fostering the reformers' belief that the substantive law needed a conception of responsibility which was sufficiently differentiated to support decisions about graduated – whether deserved or deterrent – penalties.

Intriguingly, Wiener argues that this idea of character-responsibility coincided within two strikingly different social philosophies: first, a moralised version of utilitarianism, in which the contemplation of consequences was expected to lead to more considered, rational (and, in this context, law-abiding) behaviour; second, Evangelicism, in which the contemplation of a future life was meant to have an improving effect upon self-discipline and hence character in the present one. In terms of criminal policy, these ideas entailed a view of criminal law as a key character-building institution. In this context, the early (ultimately unsuccessful) attempts to systematise criminal law through codification⁶¹ (represented by the

59 For an acute exploration of the ideas of character motivating the development of imprisonment as a form of punishment in the late eighteenth and early nineteenth centuries, and an intriguing (though not, in my view, entirely persuasive) argument about their literary origins, see John Bender, *Imagining the Penitentiary* (Chicago: University of Chicago Press, 1987).

60 One interpretation of this development is that it marks an intermediate, bridging stage between the eighteenth century's strongly character-based conception of responsibility and the later emergence of capacity as an object of legal inquiry: though assumed, the idea of capacity is introduced to the conceptual frame, facilitating later attempts to distinguish, legally, and on the basis of capacity, between the corrigible and incorrigible, occasional and habitual offender: see Wiener, n 58 above, ch 9.

61 Ironically, the movement for codification had its main success beyond the shores of England, the recommendations of rationalisation and efficiency proving more compelling in the colonial context: on the Indian Code, see Stephen, n 8 above, vol 3 ch XXXIII.

appointment of a series of Criminal Law Commissions from 1833 on) reflect at least three things: first, a confidence in the state's power to develop criminal law and punishment in this direction; second, a strong moral certainty about the propriety of enforcing certain habits and dispositions through criminal law; and third, an idea of individual responsibility which presents a curious (to the late-twentieth-century eye) mix of moralism and utilitarianism (in the sense of susceptibility to deterrence). Anyone interested in the work of Michel Foucault will doubtless detect strong Foucaultian undertones here, although Wiener's story is in some respects more finely textured than Foucault's famous account of the movement from punishing the body to punishing the soul.⁶²

Moreover – and crucially from my point of view – this early Victorian notion of responsibility as based on character had very little difficulty in accommodating either strict liability or what we would now call negligence. The vast expansion of so-called regulatory offences from the middle of the nineteenth century on – prompted largely by industrialisation and urbanisation and facilitated by the expansion of summary trial by magistrates⁶³ – is seen as entirely consistent with individual responsibility. For responsibility is based not on the direct capacity to control and understand the relevant criminal act, but rather on the defendant's assumed character or disposition. These new offences, widely interpreted by the judiciary as imposing strict liability, are probably the most striking development in substantive English criminal law of the period. They were seen, I would argue, as enforcing certain standards of care, particularly among those occupying potentially disorderly or dangerous social roles: roles which themselves generate normative standards of character – the responsible factory-owner, the decent publican – which supplemented more general understandings of respectability (or its opposite) in the expanding world of criminal law. There is an interesting analogy here with Stephen's continuing focus on conduct (as opposed to responsibility) which I noted earlier.

In the latter part of the century, however, Wiener argues that this (to our eyes) odd equilibrium of utilitarianism and Victorian moralism began to break down. The analysis is a complex one, but one of its most important components is a shift in world-view occasioned by the growing influence of the social and natural sciences, which gradually undermined confidence in individual responsibility for crime. As it became more plausible to believe that crime was – at least in some cases – the product of either social environment or physical pathology, the legitimacy of harsh, uniform punishments became more and more questionable. Conversely, the image of the criminal became a less threatening and powerful one, with many categories of offender increasingly seen as victims of mental incapacity or social deprivation and hence as proper objects not so much of punishment as of medical or welfare-oriented intervention. One of the ironies of this story is that the consequent weakening of the severity of criminal policy appears to have been facilitated precisely by the perceived success of the harsh early-Victorian approach: with crime rates falling, fear of crime was reduced, and hence became less of an obstacle to penal moderation.

Wiener's is a powerful and fascinating account of the development of criminal policy and of underlying ideas of criminality through the nineteenth century. From a legal point of view it is, however, somewhat frustrating, because he has relatively

62 Michel Foucault, *Discipline and Punish* (Harmondsworth: Penguin, tr A. Sheridan, 1979).

63 See Farmer, n 28 above, chs 3 and 4.

little to say about what was happening at the level of criminal law doctrine itself.⁶⁴ On the face of it, his story of the gradual rationalisation of criminal procedure and penal policy might be expected to have presented strong pressure for the criminal law itself to generate standards and elaborated criteria of responsibility which supported and rationalised the more systematic prosecution and, in particular, sentencing decisions being made. Yet, in fact, and notwithstanding the efforts of the would-be reformers and codifiers of criminal law, the law itself remained in a relatively unsystematic state throughout the century. In so far as Wiener has an account of the tardy rationalisation of criminal law, it is simply in terms of the inherent conservatism of the legal profession. Apart from being unduly general, this explanation sits unhappily with the fact that the profession had to – and did – adapt itself to radical procedural changes throughout the century as the adversarial trial dominated by prosecution and defence counsel and structured in terms of elaborated rules of evidence gradually emerged from an inquisitorial system dominated by judge and jury.⁶⁵

Without making any further comment on Wiener's account, let me therefore sketch another account of the same period which might be thought to promise a remedy to this deficiency – The Legal Historian's Tale. This can be exemplified by Keith Smith's *Lawyers, Legislators and Theorists*,⁶⁶ a detailed and scholarly study of the development of English criminal jurisprudence from 1800 to 1957. Unlike Wiener, Smith's analysis focuses only on legal or directly legally relevant sources: case law, legislation, the reports of the Criminal Law Commissioners, parliamentary debates on criminal law reform, and the treatises of legal philosophers whose work impinged on law reform debates (notably Jeremy Bentham and John Austin). Perhaps not surprisingly, he comes up with a rather different story.

According to Smith, what is happening during the nineteenth century is a gradual development and refinement of notions of responsibility based on individual will and traceable back in more or less unbroken sequence at least to Blackstone and other eighteenth-century commentators on criminal law such as Matthew Hale.⁶⁷ Like Wiener, Smith identifies both a strong reaction against eighteenth-century levels of discretion and the decisive influence of Utilitarian philosophy and the deterrence-based view of criminal justice which it engenders. But, unlike Wiener, Smith emphasises the fact that both Bentham and Austin insisted that, from the point of view of deterrence, there was a crucial difference to be drawn between conscious and unconscious, intended and unintended, advertent and inadvertent behaviour: the person who does not think about what she is doing, or does not advert to the relevant circumstances or consequences, is not deterrable in the same way that the advertent person is. This leads Smith to place enormous emphasis on the importance of the kind of subjective, capacity-based responsibility which predominates in late-twentieth-century criminal law theory.

Smith's thesis has undoubted power: certainly, there is evidence that the foundations for the (in my view) later development of a conception of genuinely

64 In more recent work, Wiener has begun to address this question in relation to the development of a set of *mens rea* and defence doctrines in the specific context of murder: 'Judges and Jurors: Courtroom Tensions in Murder Trials and the Law of Criminal Responsibility in Nineteenth Century England' (1999) 17 *Law and History Review* 461.

65 See Langbein, n 56 above; J.H. Langbein, 'Shaping the Eighteenth Century Criminal Trial: A view from the Ryder Sources' (1983) *University of Chicago Law Review* 1; Langbein, n 48 above.

66 Smith, n 34 above.

67 *The History of the Common Law of England* (1713) (Chicago: University of Chicago Press, 1971).

subjective criminality were being laid during the nineteenth century, and Smith provides a persuasive account of some of the intellectual and doctrinal roots of this development. Unfortunately, however, he tends to read his nineteenth-century legal evidence through his nineteenth-century intellectual or late-twentieth-century lenses. Certainly, he is far too conscientious a scholar to disguise the fact that the drift of criminal responsibility through the century was far from a linear one towards the greater adoption of subjective principles. But one feels that he is deeply puzzled as to why that was *not* the case. And, as a result of his focus on subjective responsibility, he fails to see that for nineteenth-century criminal lawyers (as opposed to commentators) the subjective/objective, advertent/inadvertent distinction did not have the same significance that it has for us today.⁶⁸ The main reason for this is something which is so pervasive that even Smith's lenses cannot filter it out: it is the simple fact that throughout the nineteenth century, and indeed well into the twentieth, English criminal law operated on the basis of a presumption that a defendant had intended the natural consequences of her actions.⁶⁹ So the analytic basis for a distinction between subjective and objective was lacking, or was developing only very slowly via a debate about how the presumption that natural consequences were intended could be rebutted. In the continued vigour of this presumption, one might argue that a modified version of Fletcher's idea of manifest criminality – and along with it a significantly character-based approach to criminal responsibility – outlived by many decades the locally based, relatively restricted criminal justice system in which it had its roots.

Implications for criminal law theory

I have only been able to give a small flavour of these two stories – each a compelling narrative in its own way. But I want, finally, to draw out of them some tentative conclusions which will inform my future research.

First, a point about method. One of the fascinating things about these two books is that one could almost read them without realising that they are about the same subject. Though one can draw out some common points – notably the importance of individual responsibility in the early part of the nineteenth century and the influence of medical science on the debate about insanity – in most respects, one could be reading about totally different worlds. The reason relates to a point I made earlier about the distinction between form and content; while Wiener's analysis is mainly focussed on the *substance* of criminal policy, Smith's is directed to the *conceptual framework* of liability – a focus which is so acute that the massive expansion in criminal law's substantive reach during the century,

⁶⁸ As witnessed, for example, by the way in which both Stephen and Kenny run objective and subjective standards together in their expositions, often blurring any possible boundary between the two: Stephen, n 8 above, vol II, 84; C.S. Kenny, *Outlines of Criminal Law* (Cambridge: Cambridge University Press, first ed, 1902) 40. In other respects, however, Stephen himself provides an earlier instance of the present-centred historical focus which I have criticised in Smith: this is particularly vivid in his interpretation of malice in early texts, such as Bracton and Foster, in terms of a nineteenth-century version of intention (see in particular vol II, 118; vol III, ch XXVI, especially 76). For a contemporary view, from the United States, see Wendell Holmes, n 29 above, Lecture II of which sets out a significantly foresight- rather than intent-based approach to criminal responsibility, but one on which the distinction between subjective and objective tests of foresight is regarded as unimportant: see in particular 44–46.

⁶⁹ See Stephen, *ibid.*, 111; Kenny, *ibid.*, 333.

along with the procedural changes which made it possible, more or less escape his comment.

On another level, the merits and limitations of the two books are complementary. Wiener's book – despite its subtitle, 'Culture, Law and Policy' – actually has remarkably little to say about developments in the judicial interpretation of criminal law. Conversely, Smith's legal focus enables him to identify, yet prevents him from exploring the significance of, a number of features of nineteenth-century criminal law which would – had he attended to them – have been of obvious relevance to Wiener's cultural analysis. In other words, Smith's legal focus raises questions which only a broader analysis can answer, while Wiener's cultural analysis ignores important legal data. Though the distinctive *modus operandi* and institutional structure of law is indeed – as its often tardy reform suggests – relatively autonomous, it is surely not isolated from broader social forces. The autonomous nature of law has been theorised by Niklas Luhmann and by Gunther Teubner⁷⁰ in terms of Luhmann's account of the social world as a set of autopoietic, self-referential systems. Yet even on this model, systems are cognitively open at the same time as they are normatively closed. It seems implausible, moreover, to suggest that developments in the form of criminal law are entirely unrelated to those in its substance.

So much for questions of method. Let me turn, finally, to one, speculative hypothesis which my research so far has suggested, and which demonstrates the potentially fruitful interaction between legal and social history. The hypothesis is premised not just on the two books I have discussed but also on a wide range of work on important developments in criminal procedure which I have not been able to discuss in this paper. These include the emerging role of counsel and of the adversarial system, with its silencing of the defendant⁷¹ and its radical reinterpretation of the respective roles of judge and jury in the light of the increasing dominance of lawyers; the gradual development of the law of evidence; the emerging role of 'expert' witnesses; the growth of systematic law reporting; and the (extraordinarily slow) construction of a system of criminal appeals.

We might hypothesise that criminal law, understood more or less as we understand it today, faces two rather general problems of co-ordination and legitimation. The first is a problem of values: how are the actors in the criminal process to co-ordinate on certain values rather than others? How are these values which criminal law expresses (and enforces), along with the substantive judgments which (purportedly) flow from them, to be legitimated? And how important is legitimation to the co-ordination of value-judgments within the criminal process? The second is a problem of knowledge: how is co-ordination on knowledge of (or belief in) the facts on which judgments in individual cases are based to be achieved? And how, when questioned, is such knowledge to be legitimated?

As far as this second problem – the problem of knowledge – is concerned, at the end of the eighteenth century, English criminal law was still in a deep sense a process based on local knowledge. Both the structure of the process, which was far more inquisitorial and focused on the pre-trial stage than today, and the jury trial itself, exemplified a confidence in collective – and often in lay – findings. Indeed, in this system the boundaries between knowledge and value co-ordination were

70 N. Luhmann, *A Sociological Theory of Law* (London: Routledge and Kegan Paul, tr B. King and M. Albrow, 1985); G. Teubner, *Law as an Autopoietic System* (Oxford: Basil Blackwell, 1993)

71 A development which, ironically, coincides with the growing belief that the defendant's interior, mental world is a proper object of investigation at the criminal trial.

blurred in the sense that this confidence extended seamlessly from fact-finding to the recognition of whether these facts constituted criminal wrongdoing. Evidence about the wide spectrum of social groups represented among those bringing prosecutions in the eighteenth century reminds us that this equilibrium depended on more than just the (undoubtedly crucial) fact that a weak central state had to – and could – rely on local elites to administer criminal justice.⁷² At the level of the problem of knowledge and shared recognition, this was still a world in which – as Fletcher puts it – crime was ‘manifest’, and when a manifestly criminal act was perceived, it made little sense to ask ‘was the person responsible for it?’⁷³ Rather, the boot was on the other foot; it was assumed that the person was responsible unless there was a special, widely recognised reason to displace that presumption.⁷⁴ This explains the gradual elaboration of defences and arguments for rebutting the inference of intended natural consequences as refinements to a system which relied heavily on character evidence to underpin the judgment of criminality made by juries which were no longer composed of witnesses but rather of evaluators of evidence, and which assigned a key role to the local resources of a lay magistracy and of active and experienced petty juries.

At the level of the problem of value-co-ordination, while – as social histories of crime and punishment in the eighteenth century show – there was a high level of conflict (especially in the fields of public order and some forms of property crime), those responsible for the administration of criminal law had confidence in the reliability of the locally based system. They assumed that, in the vast majority of cases, the strength of local understandings of criminal norms and the vitality of local enforcement institutions in co-ordinating the factual basis for the application of those norms would sustain adequate practices of conviction and punishment. And they had confidence in their power, on the relatively rare occasions when those understandings broke down and conflict erupted, ultimately to impose order from above and to negotiate conflicts as and when they arose.⁷⁵

72 On the social profile of prosecutors, see Beattie, n 56 above, 8–10 and ch 2. In his *History*, n 8 above, vol I, 354–355, Stephen makes the persuasive suggestion that the absence of a presumption of innocence in the earlier criminal law was itself a product of the weakness of the state.

73 This is also marked in the prevalence of ‘*mens rea*’ terms like ‘malice’, which are strikingly different from the ‘factualised’ conceptions of intention and recklessness which begin to emerge in the latter half of the nineteenth century (see n 49 above). The more evaluative terms such as malice connect, as I have argued, with a sense of the criminal process as a non-expert one in which lay actors can be relied upon to recognise the relevant genre of wrongdoing when they see it. On the significance of the emergence of a non-evaluative concept of ‘*mens rea*’, see also Norrie, n 2 above, Part II. The fact that the system was also underpinned by the courts’ knowledge that victims were unlikely to go to the trouble and cost of prosecution if informal social pressure could produce a satisfactory resolution of the issue further serves to explain the fact that, as late as the early nineteenth century, it would be seriously misleading to say that criminal trials were structured around a presumption of the defendant’s innocence.

74 As George Fletcher quite rightly suggests, this is still the way in which we approach conduct in criminal law today: our confidence in common-sense attributions of responsibility on the basis of appearances persists in relation to voluntary conduct just as it underpins the presumption of sanity, despite having been displaced in relation to *mens rea*.

75 As Beattie’s (n 56 above) data on prosecution show, criminal law apparently enjoyed wide legitimacy among even the relatively modest social classes; since prosecutions depended heavily on the victim’s willingness to pursue the matter to the courts, the fact that assize records show significant numbers of labourers and artisans among prosecutors suggests that, particularly in the field of ‘ordinary’ property crime and offences against the person, a wide range of social groups had, notwithstanding the class bias of the system of property which it protected (see Donald Hay, ‘Property, Authority and the Criminal Law’ in D. Hay, E. P. Thompson, P. Linebaugh and J. G. Rule (eds), *Albion’s Fatal Tree* (Harmondsworth: Penguin, 1975)), substantial faith in the criminal law. On this point see also Peter King, ‘Decision-Makers and Decision-Making in the English Criminal Law, 1750–1800’ (1984) 27 *Historical Journal* 25.

From the late eighteenth century on, the social conditions which sustained this equilibrium of knowledge and value co-ordination and legitimation were beginning to break down. With urbanisation, greater social mobility and the increase in democratic sentiments, both co-ordination and legitimation problems for criminal law began to change. Political power being markedly hierarchical in distribution and authoritarian in style, the immediate crisis might have been expected to be one of knowledge rather than of value, and in this context it is highly significant that anxieties about the efficacy of the criminal process in the late eighteenth and early nineteenth century were most acute in London, prompting the gradual professionalisation of policing and lower-level courts in the capital earlier and to a greater degree than elsewhere.⁷⁶ Yet, by and large, the heavily moralised notion of individual responsibility noted by Wiener still expressed a deep confidence in the value judgment being made by criminal conviction. Conversely, the emphasis on deterrence – on publicity and uniformity of threat – attempted to address not only legitimacy problems in a world at once increasingly fragmented and increasingly egalitarian in outlook but also new problems of co-ordinating knowledge in a mobile society. However, contrary to the assumptions of not only Smith but also many other legal historians, on my reading of the evidence, this breakdown is not yet complete at the end of the nineteenth century.⁷⁷ This raises interesting questions about the continuing multiplicity of practices of attributing responsibility in contemporary criminal law.

In this context, the more or less undiminished importance of the presumption that a person intends the natural consequences of her acts – the most uncomfortable thing about nineteenth-century criminal law from Smith's point of view – may be interpreted as having a dual significance. On the one hand, and particularly at the start of the century, it expressed a certain confidence in the propriety of judge and jury's substantive judgment: a lack of anxiety about value-co-ordination, a confidence in a certain conception of manifest criminality. But gradually it came to be seen as a way of solving the problem of knowledge co-ordination: particularly once the growth of psychological notions of individual responsibility began to pose apparently intractable problems of proof for courts and juries, the presumption provided a solution by referring to the defendant's interior, mental world yet removing the need for any close investigation of its content except where there was specific reason to rebut the presumption.⁷⁸ This argument would also explain the fact that Stephen – a legal commentator usually assured almost to the point of dogmatism – refers to the prevailing uncertainty about whether the status of the presumption: an uncertainty which suggests that it was in the process of transformation from a legal to an evidential presumption.⁷⁹ Hence a legal doctrine which found its roots in one social, procedural and legal world survived because it made sense, albeit for different reasons, in another.

76 Beattie notes that perceptions of the crime problem in London were of particular influence in driving changes in policy; fears about rising crime rates in the capital in the second half of the eighteenth century prompted the influential political classes, who were either based in London or had significant business or social interests there, to push for a reform towards a more efficient criminal justice system (Beattie, n 56 above, eg 14–15).

77 For further evidence of the continued vigour of the locally based system in the latter half of the nineteenth century, see Conley, n 56 above.

78 As Cairns, n 56 above, ch 4, esp 77–87, has argued, one significant feature of the eighteenth and early-nineteenth-century criminal process was its foundation in lay knowledge and 'common sense'; one of the most striking changes during the nineteenth century is the development of an increasingly professionalised system, and one in which expert skill and knowledge began to be seen as central to the judgment of criminal guilt.

79 Stephen, n 8 above, vol II, 111; Kenny, n 68 above, discusses the same ambiguity: see 333.

Since my research is still at an early stage, my conclusions in this paper must be tentative ones. Important questions remain to be answered about the development of an increasing need to separate form from substance in criminal law as values and processes diversify, and how this feeds into a further expansion (and mentalisation) of the *mens rea* doctrine; about the factors underlying the general trajectory in criminal law thinking from special to general part; from procedure through substance to form; from conduct through excuses to *mens rea*. There also remains an intriguing puzzle about why the criminal law in England has proved so resistant, relative to criminal procedure, to rationalisation; and about the persistence of the language of wickedness and of malice in the contemporary criminal law, where it co-exists, perhaps uncomfortably, with a psychological discourse of subjective *mens rea*. None of these questions can, needless to say, be addressed in this paper. What I do hope to have illustrated are the insights to be gained from constructing a dialogue between historical and contemporary criminal law scholarship; between scholarship on the criminal process and the criminal law; and between social and legal analyses of responsibility. In doing so, I hope also to have sketched a conception of criminal law theory as an interpretive project which has to draw not only on the standard resources of doctrinal analysis refined by philosophical argument, but also on the resources of history and of the social sciences.

Police Complaints and Criminal Prosecutions

Graham Smith*

The police complaints process is the sole means by which criminal proceedings are initiated against police officers after allegations by members of the public that they were the victim of an offence committed by officers when in the execution or purported execution of their duty. Yet this state of the law has hardly figured in recent debate, which has seen the complaints process examined almost exclusively as the preliminary stage of the disciplinary process. This paper considers police complaints, the criminal liability of the officer and the implications for reform of the process after incorporation of the European Convention on Human Rights.

Public confidence in the police is fundamental to democratic policing and satisfaction with the complaints process is essential if the police are to secure and maintain public support. Within the last decade four inquiries – by the Royal Commission on Criminal Justice,¹ the Home Affairs Committee (HAC),² the Stephen Lawrence Inquiry³ and the European Committee for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment (CPT)⁴ – have questioned whether the police complaints process enjoys public confidence. Their reports contribute to the ongoing debate on whether the police should be responsible for the investigation of complaints against their colleagues; which dates back to the 1962 Royal Commission on the Police.⁵ In May 2000 a KPMG report, commissioned by the Home Secretary,⁶ and a parallel report published by Liberty,⁷ outlined proposals for how independent investigation of complaints might operate in practice.

Although the KPMG and Liberty reports support the growing momentum for independent investigation, they both recommend the creation of an independent body with a similar organisational structure to the Police Complaints Authority (PCA).⁸ Shortly before publication of the two reports the European Court of

* Department of Social Sciences, Nottingham Trent University. Based on a paper presented to the SPTL 2000 Annual Conference Criminal Justice Panel, University College London, September 2000. Thanks to the two anonymous referees for their helpful comments.

1 *Royal Commission on Criminal Justice, Report* Cm 2263 (1993).

2 *Home Affairs Committee, First Report, 'Police Disciplinary and Complaints Procedures'* HC 258 (1998).

3 *The Stephen Lawrence Inquiry: Report of an Inquiry by Sir William Macpherson of Cluny* Cm 4262 (1999) ch 46.

4 *Report to the United Kingdom Government on the visit to the United Kingdom and the Isle of Man carried out by the European Committee for the Prevention of Torture and Inhuman or Degrading treatment or Punishment from 8 to 17 September 1997*, para 37; published January 2000, available on <www.cpt.coe.int>.

5 *Royal Commission on the Police, Final Report* Cmd 1728 (1962). As early as 1928 a minority of the Savidge Inquiry recommended that the DPP should have the power to investigate allegations against police officers, 'Inquiry in regard to the interrogation by the police of Irene Savidge under Tribunals of Inquiry (Evidence) Act 1921' Cmd 3147 (1928), *Parliamentary Papers* (1928) vol 12, 87.

6 *Feasibility of an Independent System for Investigating Complaints Against the Police* (London: Home Office Policing and Reducing Crime Unit, Police Research Series Paper No 124, 2000).

7 J. Harrison and M. Cunneen, *An Independent Police Complaints Commission* (London: Liberty, 2000).

8 Established by Police and Criminal Evidence Act 1984 (PACE) s 83 and maintained by Police Act 1996, s 66.

Human Rights (ECtHR) made a novel contribution to the debate by questioning the independence of the PCA in *Sultan Khan v UK*.⁹ Ruling that the police complaints process did not provide an effective remedy for a breach of the European Convention on Human Rights (ECHR), in violation of Article 13, the Court noted that the PCA is a body appointed by the Home Secretary and has to have regard to his guidance.

The case before the ECtHR concerned a violation of Article 8 of the ECHR, the right to privacy, where police officers had relied on intrusive surveillance to secure a conviction. It was in this context that the Court determined that the police complaints process 'does not meet the requisite standards of independence to constitute sufficient protection against the abuse of authority'.¹⁰ There is, it is argued, a deeper significance to Strasbourg's comment. In *Sultan Khan v UK* it was not alleged that a police officer had committed criminal offences; the Article 8 finding was on grounds that the surveillance methods used were not in accordance with the law. Yet, the statutory police complaints process does not distinguish between a complaint made by a member of the public according to whether it is alleged that a police officer committed a criminal offence or unsatisfactorily performed his/her duties. The PCA has the same responsibilities for complaints regardless; the only distinction is that there is a requirement for the PCA to supervise complaint investigations which are connected with death or serious injury.¹¹ The fact that 'political nominees'¹² play a part in the criminal process is damaging to the integrity of the process and raises serious questions regarding the manner in which some criminal allegations against police officers are treated at the pre-prosecution stage.

Rather than pursue the *who* investigates issue, this paper adopts a different approach to police complaints by considering the *function* of the process, and concentrates on its relevance to criminal proceedings. It is argued that while debate has prioritised who should have responsibility for managing complaints, sight has been all but lost of the accountability of the police officer for his/her wrongdoing, most particularly with regard to criminal liability. Evidence to the 1997 HAC inquiry on police discipline and complaints will be briefly examined in order to support this argument. There follows a discussion on recent developments in ECHR case law on the state's positive obligations to prevent and investigate crime, identify and prosecute offenders and provide access to the investigatory process. It is contended that with regard to police crime, incorporation of the ECHR under the Human Rights Act 1998 will have a significant impact on the police complaints process.

The function of the police complaints process

The police complaints process was first codified under section 49 of the Police Act 1964 on the recommendation of the 1962 Royal Commission on the Police. The majority of Commissioners recommended a police misconduct package consisting of three elements:

9 [2000] 8 BHRC 310. The ECtHR reiterated the finding of the European Commission on Human Rights in *Govell v UK* [1999] EHRLR 121.

10 n 9 above

11 Police Act 1996 s 70.

12 Using Sir Robert Mark's term for intended appointees to the PCA's predecessor, the Police Complaints Board, *In the Office of Constable* (London: Collins, 1975) ch 16.

- (i) the decision to bring criminal proceedings against a police officer to be made by the Director of Public Prosecutions (DPP) (enacted as section 49(3) of the Police Act 1964 and now provided for by sections 74 and 75(3) of the Police Act 1996);
- (ii) introduction of a vicarious liability rule for the tortious conduct of a police officer when acting in the performance or purported performance of the police function (chief officers were made vicariously liable under section 48(1) of the 1964 Act, now provided for by section 88(1) of the 1996 Act);
- (iii) chief officers made responsible for recording and investigating complaints against the police (enacted as section 49(1) of the 1964 Act and now generally provided for by Part IV of the 1996 Act).

The Royal Commission was of the opinion that the availability of separate procedures to deal with different types of misconduct would ensure that allegations would not be subject entirely to an internal police process:

the only clear field in which disposal of a complaint will ... be dependent on the unaided decision of a chief constable will comprise complaints either not so serious as to give rise to legal action, or in which the complainant for reasons that may appear reasonable to him does not wish for a prosecution or is unwilling to sue¹³

The Royal Commission's approach was to assume the police officer's criminal and civil liability *a priori*, allowing for less serious complaints to be resolved under an internal discipline process. This sub-division of police wrongdoing into three categories depending on whether suitable for criminal prosecution, civil action or disciplinary proceedings implies that each category has an appropriate remedy. It is difficult to conceive how remedies commensurate with the seriousness of the wrongdoing can ultimately attach unless there is some means of classifying complaints according to the seriousness of the allegation in the first instance. However, the Royal Commission did not recommend a classification system and section 49 of the Police Act 1964 did not define a complaint, an omission which allowed for all allegations to be recorded and investigated in the same manner regardless of whether an officer had been accused of a criminal offence or incivility.¹⁴

There was an inherent danger in the Royal Commission's proposals for the substitution of the criminal process with administrative procedure, as was recognised by R. E. Winterbottom MP during the committee stage of the Police Bill:

I hope complaints are going to be dealt with purely and simply as complaints that have been raised against the police in the exercise of their duty, and not as another means of investigating crimes that have been committed.¹⁵

It is suggested that this is precisely what has happened since codification of the complaints process. A consequence of section 49 of the Police Act 1964, and preserved by subsequent reform under PACE and the Police Act 1996, is that the complaints process serves as the sole means by which criminal proceedings are initiated after a member of the public alleges that s/he has been the victim of a criminal offence committed by a police officer when in the execution or purported execution of his/her duty.

¹³ n 5 above, 138.

¹⁴ Unless the complaint alleged an offence with which the police officer had been charged. When a 'complaint' was eventually defined in PACE, s 84(4) it was in general terms as 'any complaint about the conduct of a police officer', a definition retained under the Police Act 1996, s 65.

¹⁵ HC Standing Committee D col 725 13 February 1963.

Police legitimacy, underpinned at common law by the office of constable,¹⁶ requires adherence to the principle that the police officer should be accountable to the law for his/her criminal wrongdoing in like manner as the citizen. On analysis, the 'like-manner' principle can be separated into two limbs – first, that responsibility for the process, or stages of the process, must rest with the same authority and secondly, that the same process must apply. The conventional approach to the principle has been insistence, articulated primarily by police organisations, that a criminal allegation against a police officer must be investigated by the same institution as would investigate an allegation against a member of the public – the police. This has been relied on as grounds for resisting any form of independent involvement in the investigation of complaints, most notably by former Commissioner of Police of the Metropolis (CPM) Sir Robert Mark.¹⁷ Whereas investigation of police officers by their colleagues conforms to the first limb of the like-manner principle, the principle is not adhered to in that the pre-trial criminal process does not comply with the second. That is because investigation and charge of a police officer for an alleged criminal offence following a complaint is currently subject to different procedures to those that apply to the general public. It is suggested that existing criminal and police administrative law and police practice combine to allow this state of affairs.

A result of the general classification of complaints is that the reporting to the police by a member of the public of what would be a recordable criminal offence if made against another member of the public is recorded as a complaint when made against a police officer. At the commencement of the process an officer alleged to have committed a criminal offence is treated differently from the citizen with the consequence that the police officer arguably enjoys a privileged position relative to the citizen at the outset. This is because under section 65 of the Police Act 1996 the offence must be reported by the victim and not by a witness; thus the 'good citizen' is prevented from independently reporting police crime of this nature.¹⁸ Moreover, the allegation is dealt with separately from other criminal allegations in accordance with the police complaints process.

The officer concerned will be informed that a complaint has been made and that his/her conduct is the subject of an investigation.¹⁹ The ensuing investigatory process has a duality in order to cover the eventuality that criminal or disciplinary proceedings may follow,²⁰ which accounts for the different manner in which the Police and Criminal Evidence Act 1984 (PACE) provisions are applied to the police officer in comparison to the citizen. Under the Police (Conduct) Regulations 1999 the officer is entitled to the safeguards at interview which applied to the criminal suspect prior to sections 34–37 of the Criminal Justice and Public Order Act 1994.²¹ Therefore, at interview, an officer is entitled to remain silent without fear of prejudice in reply to questions regarding prospective disciplinary

¹⁶ *Fisher v Oldham Corporation* [1936] 2 KB 364.

¹⁷ n 12 above.

¹⁸ An incident was reported to the author in 1998 by a solicitor after his client was subjected to an assault in a public place following an argument, and a bystander called the police. On the police's arrival the assailant showed his police warrant card and his client was asked if he wished to make a complaint. He objected on grounds that his attacker should be arrested and charged in the same manner as if he had been the aggressor. No further action was taken after the client received a written communication from the police that the assault allegation would only be pursued if he made a formal complaint.

¹⁹ Police (Conduct) Regulations 1999, SI 1999 No 730, reg 9.

²⁰ See below on the duality of the criminal and disciplinary investigations in evidence to the HAC.

²¹ n 19 above, reg 9(e).

proceedings, whereas adverse inferences may be drawn from failure to answer the same questions in connection with criminal proceedings.

It is possible that a police officer may be arrested under statute or common law by another officer on suspicion of having committed an offence after an allegation has been made by a member of the public, as expressly provided by the complaints process.²² However, the discretionary nature of police powers to arrest on grounds of reasonable suspicion,²³ which has been increasingly interpreted by the courts in a manner which protects officers,²⁴ also serves as a safeguard for an officer accused of criminal wrongdoing. Police officers' working assumptions are not that colleagues commit criminal offences when performing onerous duties, particularly following allegations made by members of the public,²⁵ and the authorities clearly allow for the non-exercise of their discretionary powers. It is incumbent on officers to record complaints, and the most probable source of evidence that an officer has committed an offence is at interview. It is suggested that the same working assumptions that prevent officers from arresting colleagues also characterise interviews conducted as part of complaint investigations, and that the right-of-silence safeguards enjoyed by officers in disciplinary matters extends to protect them against criminal proceedings. Where a complaint is made and the complainant faces criminal proceedings in connection with that complaint, a *sub judice* rule applies which delays investigation of the substantive complaint until after the conclusion of those criminal proceedings. It is apparent that the operation of this rule serves as further protection for an officer, particularly if the officer arrested the suspect and caused him/her to be charged in order to conceal his/her own criminality.²⁶ Finally, with regard to the existence of different procedures for police officers, whereas a member of the public suspected of an offence will normally be charged by a custody officer under section 37(7) of PACE, the decision to initiate proceedings against police officers is made by the DPP under sections 74 or 75(3) of the Police Act 1996.²⁷

Having argued on grounds of principle that a combination of legal procedure and police practice has contrived to create a different pre-trial criminal process for police officers, it is now necessary to add substance to the argument that administrative procedure has replaced due legal process. Police complaint statistics are 'confusing and misleading',²⁸ and the practice of regularly changing the way they are compiled and presented, by Her Majesty's Chief Inspector of Constabulary (HMCIC), the CPM and the PCA, inhibits analysis of long-term trends. Nevertheless, a constant feature of published statistics has been the infrequency of criminal proceedings against police officers. Between 1969 and 1989 consistent figures were published in HMCIC's annual reports revealing that a mean average of just over 30 officers were prosecuted annually for non-traffic

22 n 14 above.

23 *Holgate-Mohammed v Duke* [1984] 1 AC 437.

24 *Castorina v Chief Constable of Surrey* (1988) New LJ 180.

25 S. Box and K. Russell, 'The Politics of Discreditability: Disarming Complaints Against the Police' (1975) 23 *Sociological Review* 2; A. Sanders, 'Controlling the Discretion of the Individual Officer' in R. Reiner and S. Spencer (eds), *Accountable Policing* (London: Institute for Public Policy Research 1993).

26 K. Russell, *Complaints against the police which are withdrawn* (Leicester: Polytechnic Reprographic Unit, 1986), suggests officers use this as a common ploy to evade the consequences of their conduct. Supported by the author in view of the high incidence of false imprisonment and malicious prosecution claims in successful actions against the police; G. Smith, *Police Crime: A Constitutional Perspective* (unpublished doctoral thesis, 1998) ch 2.

27 s 75(3) is the standard procedure and s 74 is for senior officers.

28 M. Maguire and C. Corbett, *A Study of the Police Complaints System* (London: HMSO, 1991) 35.

Table 1: Police officers in England and Wales convicted of criminal offence(s): 1992–1998/9

	1992	1993	1994	1995/6	1996/7	1997/8	1998/9
Officers convicted of criminal offence(s)	289	289	226	260	222	218	240
Officers convicted of non-traffic offence(s)	57	49	35	45	43	53	65
Officers who resigned before criminal proceedings concluded	n/a	19	12	22	22	12	15
Criminal proceedings arising from substantiated complaints						23	38
Officers convicted of non-traffic offence(s) arising from a complaint						16	18

Source: Home Office Statistical Bulletins, *Police Complaints and Discipline in England and Wales*, 1993–1998/9.

offences arising from complaints in England and Wales (excluding the Metropolitan Police Service (MPS)).²⁹ Between 1983 and 1989 a total of 46, giving an annual mean average of just under eight, Metropolitan officers were convicted of non-traffic offences arising from complaints.³⁰ Combined figures for police forces throughout England and Wales, are available from the Home Office since 1992, as shown in Table 1.

In order to gain a full impression of the situation, it helps to look at the numbers of complaints recorded, substantiated and the disciplinary outcomes, this time since 1994 as consistent figures only date back that far.

A striking feature of the statistics is that only 20 per cent of substantiated complaints lead to criminal or disciplinary proceedings. The reason for this is that the

Table 2: Police complaint statistics for England and Wales: 1994–1998/9

	1994	1995/6	1996/7	1997/8	1998/9
Total complaints recorded	36,521	35,840	36,731	35,820	31,653
Complaints investigated	9,590	8,653	10,820	9,832	9,202
Complaints substantiated	793	749	834	847	745
Disciplinary proceedings arising from substantiated complaints	133	167	148	172	141
Officers convicted of disciplinary offence(s) arising from a complaint	96	117	102	113	116

Source: Home Office Statistical Bulletins, *Police Complaints and Discipline in England and Wales*, 1994–1998/9.

²⁹ Annual reports of Her Majesty's Chief Inspector of Constabulary, 1969–1989.

³⁰ Annual reports of the Commissioner of Police of the Metropolis, 1983–1989.

majority are considered not to require disciplinary action and are followed by a supervisory officer giving words of advice or an admonishment to the officer complained against. However, given the disparity between the number of successful malfeasance claimants (see below) and officers who face criminal or disciplinary proceedings as a consequence of substantiated complaints, serious questions arise regarding the accountability of police officers. It is not only the officer's liability that is at issue here; so too is the manner in which the complainant is treated. To illustrate this point comparison can be made between the informal resolution process³¹ and what occurs when disciplinary proceedings do not follow a substantiated complaint. A feature of informal resolution, which is acknowledged to operate to the benefit of police and complainant,³² is that a meeting may be arranged between the complainant and officer complained against subject to the consent of both parties. This presents the complainant with an opportunity to achieve some form of satisfaction that his/her grievance has been addressed in a meaningful way, irrespective of whether or not the officer apologises for his/her action. A similar opportunity is not available to the majority of complainants who have their complaints substantiated. For these successful complainants, who participate in what may be an unpleasant experience, particularly at interview, their ordeal is concluded when they are informed in writing by the PCA that disciplinary proceedings will not follow as it has been decided that a supervisory officer will speak to the officer concerned.

The impression here is of a complaint process operating as an instrument of police management. Even where complaints have been substantiated, under 3 per cent annually of the total number of recorded complaints, it is apparent that the complaint process operates primarily to identify malpractice in the interest of police administration with scant regard for vindication of the complainant.³³

It is not suggested that low prosecution and conviction rates amount to conclusive evidence that police managers prefer to invoke the disciplinary process or that officers enjoy immunity from criminal prosecution. Yet a major cause for concern arises when criminal and disciplinary statistics are compared with the number of successful actions for damages against chief officers, as the HAC³⁴ and the CPT³⁵ point out. Again, the presentation of police action statistics has been haphazard. Since 1991 the CPM has provided the only regular figures, and HMCIC presented the national picture for the first time in his 1998/99 *Annual Report*. Throughout the 1990s there was a steady increase in successful claims against the MPS, more than doubling from 127 in 1991/92 to 295 in 1998/99, and the amount paid out in awards and settlements more than trebled from £581,000 to £2,309,000 in the same period.³⁶ In 1998/99, 1,302 malfeasance claims, covering the intentional torts of assault, false imprisonment and malicious prosecution, were settled and 65 claimants awarded damages, attracting payouts (excluding costs) of close to £4.5 million throughout England and Wales.³⁷

31 Introduced by PACE, s 85(4) and now under the Police Act 1996, s 69(4).

32 Maguire and Corbett, n 28 above, 194.

33 See G. Smith, 'The police complaint process: a means for identifying rotten apples or satisfying grievances?' (2000) 99 *The Criminal Lawyer* 2-4.

34 n 2 above, para 32.

35 n 4 above.

36 Annual reports of the Commissioner of Police of the Metropolis, 1991/92 and 1998/99.

37 *Her Majesty's Chief Inspector of Constabulary, Annual Report of 1998/99* HC 804 (1999). With police forces receiving 5,961 malfeasance claims in 1998/99 it may be that the national trend will follow the MPS pattern.

From the statistics it can be safely concluded that a person who suffers as a result of police misconduct has a far greater chance of achieving a remedy in the civil courts than by invoking the complaint process. Moreover, the intentional torts of assault and false imprisonment are also common law criminal offences and malicious prosecution closely translates to conspiracy to pervert the course of justice. Thus, the total of 18 police officers convicted of criminal offences arising out of complaints can be compared, certainly with the 65 court awards to claimants, and at a stretch with many of the 1,367 successful actions, notwithstanding that different burdens of proof apply and some cases will have been settled on a commercial basis with liability denied.

The importance of this comparative statistical exercise is not to engage in argument over the relative merits of alternative remedies to police misconduct; it is to demonstrate that police unlawfulness is more widespread than the criminal statistics suggest. It is apparent that instead of three remedies to police misconduct, as envisaged by the 1962 Royal Commission, there are two – complaints, which are unlikely to lead to meaningful action from the perspective of the complainant, and actions for damages. With hindsight, given the way the legislation was framed it was inevitable that complaints and actions would emerge as competing remedies, but of greater concern is that the criminal liability of the police officer has been largely overlooked, a development which Winterbottom rather prophetically warned against in 1964.

It is suggested that a significant reason for this oversight is that since the 1962 Royal Commission, when a minority report recommended that a Police Ombudsman or Commissioner of Rights should oversee police investigations of complaints,³⁸ debate has been dominated by the appropriateness of lay involvement in the complaints process.

An independent element was introduced to the complaint process with creation of the Police Complaints Board (PCB) under section 1 of the Police Act 1976. However, its responsibility for examining completed police investigations and power to direct a chief officer to prefer disciplinary charges³⁹ were not intended to detract from the essentially managerial function of the process. Police staff organisations and their representatives insisted on the introduction of a 'double jeopardy' rule created under section 11 of the Act.⁴⁰ Designed to protect officers from disciplinary action where criminal proceedings failed, the rule also served to deter the PCB from interfering in police affairs when acting on behalf of complainants.⁴¹

In its first *Annual Report* the PCB sought to allay police opposition by announcing it would approach its responsibilities with due regard to the difficulties faced by officers when performing their duties:

The case for action against a police officer for breach of a provision of the discipline code may have to be balanced against consideration of force morale and the effect of the decision on the police in general.⁴²

It is contended that this comment illustrates that the primary purpose of external scrutiny of police investigations and discipline decisions was to improve the management of complaints rather than to address complainants' grievances.

38 n 5 above 138.

39 Police Act 1976, ss 2 and 3, respectively.

40 See L. Leigh, 'The Police Act 1976' (1977) 4 *British Journal of Law and Society* 115.

41 G. Smith, 'Double Trouble' (1999) 149 *New LJ* 6900: 1223.

42 *Police Complaints Board, Report for the Year 1978* HC 5 (1979) 14.

The PCA replaced the PCB in June 1985, with additional powers to supervise police investigations under Part IX of PACE. Lord Scarman is widely acknowledged as the principal architect of the current process after he urged the need for greater independence to prevent loss of public confidence in the wake of the serious street disorders in 1981.⁴³ However, the supervision of complaint investigations by PCA members has not had a positive effect on public opinion as revealed by a series of surveys conducted for the Authority between 1992 and 1996.⁴⁴

The 1993 Report of the Royal Commission on Criminal Justice doubted that the complaint and discipline processes then in place commanded public confidence and described them as 'ineffective', 'lengthy' and 'uncertain'.⁴⁵ The First Report of the 1997 HAC under the new Labour administration went further by declaring that failure to improve procedures would have an adverse effect on public confidence 'both in the system and the police as a whole'.⁴⁶

To conclude this brief outline of the function of the police complaints process, reference is made to a comment made by the 1962 Royal Commission on the Police when reflecting on the character of the police discipline process prior to codification of the complaints process:

Our impression is that the Police (Discipline) Regulations, and the procedure they embody, are soundly conceived; but it is evident that the regulations have been drafted primarily from the point of view of police discipline, and only incidentally as a means of dealing with complaints against the police by members of the public.⁴⁷

The most detailed examination of the complaints process in recent times was conducted by the HAC in 1997, and culminated with its report, *Police Disciplinary and Complaints Procedures*. It is held, as the report title implies, that over 35 years after a Royal Commission attempted to address this problem the complaints process continues to play second fiddle to the needs of police discipline.

The Home Affairs Committee inquiry into police discipline and complaints⁴⁸

Two months after the election of a Labour administration in May 1997 the Chief Constable of the West Midlands Police, Edward Crew, featured on the front page of the *Independent* exclaiming that some of his colleagues could not hold down a job with Sainsbury's.⁴⁹ His outburst coincided with the election of Chris Mullin MP, no stranger to police misconduct,⁵⁰ as chair of the HAC and an announcement that the Committee's first inquiry would be into police discipline and complaints.

43 *Report on the Brixton Disorders of 10-12 April 1981* Cmnd 8427 (1981) para 8.61.

44 In the PCA's 1996 survey only 37 per cent of respondents thought it impartial with the same proportion in favour of the police investigating themselves; Police Complaints Authority, *Annual Report of 1995/6* HC 469 (1996) 48. In the only published research on PCA supervision, Maguire and Corbett found that PCA members were professional and efficient yet significantly hindered by inadequate resources; n 28 above, 197.

45 n 1 above, 48.

46 n 2 above, xvii.

47 n 5 above, para 447.

48 This section is based on a paper presented to a conference of the British Section of the European Group for the Study of Deviance and Social Control; University of Wales Bangor, April 2000.

49 10 July 1997.

50 *Error of Judgment* (Dublin: Poolbeg, 1987).

Two other developments in July 1997 were of major significance for police misconduct. On consecutive days in the Divisional Court the DPP, Dame Barbara Mills, suffered acute embarrassment as she admitted the Crown Prosecution Service (CPS) failed to properly consider whether Metropolitan police officers should face manslaughter proceedings following the deaths of Oluwashijibomi Lapite and Richard O'Brien in custody.⁵¹ As a consequence of these revelations the DPP relinquished sole responsibility for deciding whether police or prison officers should be prosecuted in serious cases and she appointed retired circuit judge, Gerald Butler, to inquire into CPS decision making procedures in such cases.⁵² There was further disappointment for the DPP on the last day of the same month when the Divisional Court ordered her to reconsider her decision not to prosecute West Midlands officers in *R v DPP ex parte Treadaway*.⁵³

On the same day that the Divisional Court gave its judgment in *ex parte Treadaway*, Sir William Macpherson was appointed by the Home Secretary to chair an inquiry into the death of Stephen Lawrence under section 49 of the Police Act 1996. The Lawrence family and their supporters' criticisms of the police response to Stephen's murder had been consistently rebuffed by the MPS and calls for a public inquiry rejected by the Conservative government.⁵⁴

This provides the context in which the HAC heard evidence on police discipline and complaints, firstly by written submissions in the late summer of 1997 and then orally between 13 October and 4 December. Even from this cursory outline of events it is apparent that there were two distinct areas of concern. On the one hand, the police establishment (consisting of ACPO, HMCIC, CPM, PCA and the Home Office) disturbed by growing evidence of corruption, sought an increase in chief officers' disciplinary powers at the expense of the safeguards, or perhaps more appropriately privileges, enjoyed by junior officers. On the other hand, the issues raised in law – in the form of actions for damages, challenges to prosecutorial decision making and the inquiries conducted by Macpherson and Butler – concerned issues arising within policed communities with the degree of accountability exercised over the police. Examination of evidence to the HAC, its recommendations to the Home Secretary and the subsequent reforms reveals that whereas chief officers' concerns have been addressed, those emerging from within policed communities have been largely overlooked to date.

The HAC published 300 pages of written and oral evidence as part of its inquiry into police discipline and complaints and the discussion below is restricted to two areas – the police misconduct problem as presented to the HAC by chief officers and the relation between criminal and disciplinary proceedings.⁵⁵

Sir Paul Condon, CPM between January 1993 and January 2000, expressed his concerns with police misconduct exclusively in terms of police officers who turn to crime for financial gain.

⁵¹ *Guardian*, 24 and 25 July 1997, respectively.

⁵² G. Smith, 'The DPP and prosecutions of police officers' (1997) 147 New LJ 6804: 1180, and 'The Butler Report: an opportunity missed?' (1999) 149 New LJ 6902: 1277.

⁵³ (1997) *Times*, 31 October. Derek Treadaway was awarded £50,000 damages for assault by West Midlands police officers after a High Court judge found he had a plastic bag placed over his head when questioned during an armed robbery investigation, *Treadaway v Chief Constable of West Midlands Police* (1994) *Independent* 23 September.

⁵⁴ For a brief chronology up to the creation of the Inquiry see Police Complaints Authority, *Report on the investigation of a complaint against the Metropolitan Police Service by Mr N and Mrs D Lawrence* Cm 3822 (1997) 2. For Mr and Mrs Lawrence's views see the *Stephen Lawrence Inquiry*, n 3 above, ch 4.

⁵⁵ Relying on the Minutes of Evidence n 2 above, vol II 1–128.

The sort of people I am speculating about are experienced, cunning, devious criminals who operate as police officers, who I think have been lost along the way. Experience suggests that they came into the service well motivated, they have operated as very good police officers, at a superficial level they are still operating as very good police officers in respect of some of the things they do. Sadly, they have given in to the opportunity to make significant sums of money through their malpractice.⁵⁶

Presentation of police criminals as wayward souls conforms with the orthodox view that corruption is a management problem – to be prevented by tight supervision and punished by strict discipline.⁵⁷ The Commissioner did not find it necessary to refer to allegations of wrongdoing made against police officers while purportedly acting in the execution of their duty by members of the public. This was all the more disturbing as the MPS had been criticised following the Lapite, O'Brien⁵⁸ and Lawrence deaths. Furthermore, members of the HAC had received written and oral evidence from Liberty and lawyers representing claimants and complainants against the police.⁵⁹ They had been informed by two experienced police litigation solicitors, who had concluded 60–70 successful actions for damages, that only one criminal prosecution had been brought in connection with their clients, one officer disciplined (by reprimand and caution) and four complaints substantiated.⁶⁰

The ACPO delegation, consisting of Peter Bensley, Chief Constable of Lincolnshire Police and Chairman of the ACPO Discipline and Complaints Sub-Committee and Paul Whitehouse, Chief Constable of Sussex Police and Chairman of the ACPO Personnel and Training Committee, also placed major stress on corruption. However, in answering questions on police actions and complaints they devoted some time to community concerns with complaints and discipline.⁶¹

Turning now to the relation between criminal and discipline proceedings the Committee acknowledged that it did not seek to cover 'the detailed interrelationship between criminal and disciplinary procedures'.⁶² This was a surprising omission given the amount of evidence devoted to the close connection, Committee members' confusion with the precise nature of the relation and the concern expressed with witnesses' responses to questions regarding prosecutions of police officers.⁶³

In written evidence the Chief Constable of West Midlands Police, Edward Crew, stressed 'The employment and criminal processes are inextricably linked in a unique way'.⁶⁴ In contrast, at an early stage in oral evidence the Chief Constable of Lincolnshire Police urged 'it is very easy to slip from criminal inquiries into personnel management procedures and I implore you to keep that dichotomy very clear in your minds throughout the whole of this process [the inquiry]'.⁶⁵ It would

⁵⁶ *ibid* 121.

⁵⁷ T. Newburn, *Understanding and preventing police corruption: lessons from the literature*, (London: Home Office Policing and Reducing Crime Unit, Police Research Series Paper No 110, 1999).

⁵⁸ Papers on these two cases had been provided to the HAC as appendices to written evidence presented by B. M. Bimberg & Co Solicitors, n 2 above, vol II 231 (not published in the Minutes of Evidence and Appendices).

⁵⁹ *ibid* 34–47 and 217–234.

⁶⁰ *ibid* 45.

⁶¹ *ibid* 1–20.

⁶² *ibid* vol I para 3 n 7.

⁶³ The DPP, Dame Barbara Mills, was robustly examined on the failures of the CPS to prosecute police officers following the Lapite, O'Brien and Treadaway challenges in the Divisional Court and to secure a conviction against a single member of the West Midlands Serious Crime Squad in the wake of 30 convictions quashed by the Court of Appeal, *ibid* vol. II 107–116.

⁶⁴ *ibid* 253.

⁶⁵ *ibid* 4.

appear that the Committee followed the latter advice. Although written evidence from the Home Office and the CPS stated the connection between criminal and disciplinary proceedings⁶⁶ and members of the HAC acknowledged the complex relation, they did not make a determined effort during public hearings to deconstruct it with the assistance of witnesses. Representatives of the Police Superintendents' Association and the Police Federation made several references to the close association, with the Federation Chairman and Vice-Chairman describing the relation as an 'overlap' on two occasions without intervention from the Committee.⁶⁷

Presenting ACPO's evidence to the Committee, Bensley focused on corruption and the powerlessness of senior officers to manage with the disciplinary powers available to them. HAC member Robin Corbett objected to this line of argument and there was the following exchange:

Corruption is not a disciplinary matter; it is a criminal offence. Why should it get anywhere near the disciplinary procedure or personnel and management?

Mr Bensley: If the inquiry into the corruption is done thoroughly, the criminal offence, separating that out completely, as I have said, is dealt with in the way that all other criminal offences are dealt with – and there is no evidence I have heard to demonstrate that it is not dealt with that way – then it will go through the due process of law, through the criminal courts and people will be dealt with accordingly.

Mr Corbett: That is not quite the end of the story, is it, because Sir Paul Condon and Mr Crew in circumstances have said 'Where we do not get enough evidence to lay criminal charges in the hope that the CPS will take those forward, we then want to fall back into this disciplinary procedure.' This is where I find confusion.

Mr Bensley: ... If you have a case and you do a criminal investigation and find that the evidence does not come up to the standard required for a criminal prosecution the Crown Prosecution Service will make a decision that no further action will be taken with that ... what you often do, quite properly, is you look through all the circumstances and you find that officers have made procedural errors ... then you can couch discipline offences.⁶⁸

Bensley went on to argue for a reduction in the burden of proof in disciplinary proceedings from the criminal standard to the balance of probabilities as a solution to this difficulty.

The Chief Constable of Sussex provided an insight into ACPO thinking on the relation between discipline and criminal proceedings in answer to a question on the ACPO proposal for a fast-track disciplinary procedure. Considering a hypothetical case where an officer is alleged to have used unreasonable force and is dismissed prior to the conclusion of criminal proceedings he explained:

There is then a public interest decision on whether it is still necessary to prosecute that police officer because the position in which he or she exercised their power is no longer occupied by them ... [W]e go from the position of having to show that people are prosecuted, which is where we are currently in order to show that justice has been done, by saying: 'We have rectified the wrong which is we have removed the individual who was misbehaving, do we now still need to prosecute?' ... On the other hand, if it is the lesser of two evils, I prefer to have the officer convicted in disciplinary proceedings and out of the service rather than acquitted in criminal proceedings and still in the service.⁶⁹

Unlike his ACPO colleagues, the Metropolitan Commissioner did not attempt to accommodate his appeal for increased disciplinary powers with the due process of

⁶⁶ *ibid* 130 and 200.

⁶⁷ *ibid* 76 and 83.

⁶⁸ *ibid* 4.

⁶⁹ *ibid* 6.

law, and made several references to Sir Robert Mark's anti-corruption purges in the 1970s⁷⁰ in support of his argument:

the knowledge and the survival cunning of criminals who serve as police officers actually makes it very difficult for the criminal system to deal with them ... he [Mark] had no difficulty in formulating discipline charges involving suspension from duty and eventual dismissal of officers in cases where the DPP had declined to prosecute or in which a prosecution had been followed by acquittal. In essence no double jeopardy and the discipline system enabled him to get at bad officers even if the criminal system had failed.⁷¹

Condon went on: 'it is now actually harder to deal with bad officers through discipline than through criminal law and that cannot be in the public interest.'⁷² This is a remarkable statement from the highest law enforcement officer in the country, and moments later Condon made some amends for his blunder when he volunteered:

Ideally and primarily we would wish to deal with these officers through the criminal law: that is what it is designed for. If we cannot do that, then we should take away the opportunity for these people to remain as police officers and utilise all their police powers, which are awesome, for their own means rather than for the public good.⁷³

It is evident that during the course of the HAC inquiry ACPO assumed a corporate stance which manifested itself in several ways. Firstly, chief officers prioritised their organisational needs by giving primacy to their responsibilities as police managers before their duties as law enforcement officers. Condon detailed his approach during an exchange on the need for a fast-track discipline procedure to dismiss officers for gross misconduct:

Mr Malins: You would be quite sure about the facts before you dismiss.

Sir Paul Condon: Yes.

Mr Malins: You are very happy in considering the position under fast track to have the standard of proof in your own mind of the criminal standard, namely you are sure.

Sir Paul Condon: No.

Mr Malins: Why not?

Sir Paul Condon: They are separate processes. This is an administrative process. This is me as an employer saying to an employee, this behaviour is so outrageous, so unsuitable to being a police officer that I am dispensing with your services at this point.⁷⁴

Condon is at fault here for representing himself as the police officer's employer. He is not.⁷⁵ Under the office of constable police officers exercise an original responsibility as law officers of the Crown.⁷⁶ In their written evidence ACPO was, again, more circumspect and suggested:

Public faith in the office of constable goes to the heart of society's trust in the criminal justice system. To enhance that faith and maintain the unique constitutional position of a constable we consider that the spirit of employment law should now apply to members of the police service.⁷⁷

Secondly, as representatives of public services chief officers have responsibility for protecting the reputation of the police. A corollary of enhanced disciplinary powers

⁷⁰ n 12 above.

⁷¹ n 2 above, vol II 117.

⁷² *Ibid* 118.

⁷³ *Ibid* 119.

⁷⁴ *Ibid* 127.

⁷⁵ *R v Chief Constable of the Devon and Cornwall Constabulary, ex p Hay* [1996] 2 All ER 711.

⁷⁶ *Fisher v Oldham Corporation* [1936] 2 KB 364.

⁷⁷ n 2 above, vol II 154.

is that chief officers have the option of quietly shedding criminal elements within their forces away from the glare of courtroom publicity. Thirdly, while not wishing to undervalue the threat to society posed by corrupt police officers, offences of this type offend primarily against the institutional norms, values and ethos of the police service; they do not address community concerns with police crime.

The HAC published its recommendations to the Home Secretary in its January 1998 Report,⁷⁸ which he broadly accepted on 23 March 1998.⁷⁹ The major changes to the discipline process introduced on 1 April 1999 were:

- Introduction of police efficiency regulations and an unsatisfactory performance process.⁸⁰
- Introduction of a fast-track disciplinary procedure.⁸¹
- Abolition of the double-jeopardy rule.⁸²
- Introduction of the civil standard of proof at disciplinary hearings (including fast track).⁸³
- Disciplinary hearings may proceed in the absence of the officer concerned.⁸⁴
- All disciplinary hearings to be heard by three police officers.⁸⁵
- In the event of a sanction imposed on the officer, the right of review to his or her chief officer.⁸⁶
- The right of appeal to a Police Appeals Tribunal set up by the Police Authority.⁸⁷

The general effect of the new regulations has been to replace the quasi-military police discipline process with employment-style labour relations practices, as urged by ACPO in its memorandum of evidence to the HAC.⁸⁸ Associated with this new approach there has been a change in vocabulary with the Discipline Code⁸⁹ replaced by a Code of Conduct⁹⁰ and officers are no longer charged with disciplinary offences⁹¹ but failure to meet the appropriate standard as outlined in the Code.⁹² In order to conform with Part IV of the Police Act 1996 and the interpretation of disciplinary proceedings under section 65, the term 'disciplinary proceedings' has been retained to describe the entire process.⁹³ However, throughout the Home Office guidelines on the new regulations the term 'misconduct' has replaced 'discipline',⁹⁴ and the term 'unsatisfactory performance procedure'⁹⁵ introduced along with 'inefficiency hearings'⁹⁶ in the efficiency regulations.

⁷⁸ *ibid* vol I para 185.

⁷⁹ HC Deb vol 309 cols 21–24 23 March 1998.

⁸⁰ Police (Efficiency) Regulations 1999, SI 1999 No.732.

⁸¹ Police (Conduct) Regulations 1999, SI 1999 No. 730, reg 39.

⁸² Police and Magistrates' Courts Act 1994, s 37(f), Police Act 1996 (Commencement and Transitional Provisions) Order 1999, S.I. 1999 No. 533 (C.14).

⁸³ n 81 above reg 23(3)(b).

⁸⁴ *ibid* reg 24(2).

⁸⁵ *ibid* reg 18.

⁸⁶ *ibid* reg 34.

⁸⁷ The (Police Appeals) Tribunal Rules 1999, SI 1999 No. 818.

⁸⁸ n 2 above, vol II 154.

⁸⁹ Police (Discipline) Regulations 1985, SI 1985 No 518, sched 1.

⁹⁰ Police (Conduct) Regulations 1999, SI 1999 No 730, sched 1.

⁹¹ n 89 above, reg 4.

⁹² n 90 above, reg 4.

⁹³ *ibid* reg 11(5).

⁹⁴ With one exception, Home Office, *Guidance on Police Unsatisfactory Performance, Complaints and Misconduct Procedures* (1999) s 3, 5; available on <www.homeoffice.gov.uk>.

⁹⁵ *ibid* s 1.

⁹⁶ n 80 above, reg 12(2).

With regard to the complaints process the HAC concluded in principle that independent investigation of complaints is desirable and recommended that the Home Secretary commission a feasibility study.⁹⁷ Whereas the Home Secretary acted relatively quickly in response to the HAC's recommendations on discipline, it was not until more than one year later when the Lawrence Inquiry made the same proposal⁹⁸ that research was commissioned.⁹⁹ At the time of writing, after publication of reports by KPMG and Liberty,¹⁰⁰ the Home Office is considering responses to its consultation paper, *Complaints Against the Police*, with the declared intention of introducing a stronger independent element and greater openness to the investigation of police complaints.¹⁰¹

Delayed reform of the complaint process means that the 1999 police conduct regulations, along with the Home Office guidelines,¹⁰² have to be read with the 1985 complaints regulations.¹⁰³ Although one section of the guidelines is devoted to complaints procedures,¹⁰⁴ the misconduct guidelines refer to complaints as one of the means by which information about an officer may come to the attention of police managers which requires investigation as part of the misconduct process.¹⁰⁵ Acknowledging that there must have been some difficulty drafting the guidelines, given the need to reconcile the reformed character of the new regulations with the old, it is clear that the complaints process is presented as an element of the misconduct process, and regarded as a tool of police management.

The European dimension

It took nearly six years for the police discipline reforms, including abolition of the double-jeopardy rule, recommended by the Royal Commission on Criminal Justice to be put into effect; five years after enactment by the Police and Magistrates' Courts Act 1994. At the same rate, a more independent and open complaints process might be implemented some time in 2005, at the earliest! On the evidence presented to the HAC it is apparent that ACPO, HMCIC and PCA support for transformation of the discipline process speeded up the pace of change. In contrast, in their evidence to the HAC, representatives of the police establishment did not give the impression that they were unduly worried with concerns raised in policed communities regarding the accountability of police officers in response to complaints made by members of the public. It is also apparent that chief police officers are uncomfortable with independent and open investigation of complaints,¹⁰⁶ which will be a major obstacle

97 n 2 above, para 81

98 n 3 above, recommendation 58

99 n 6 above, iii.

100 n 7 above.

101 Home Office *Complaints Against the Police: A Consultation Paper* (London: HMSO, 2000) para 6.

102 n 94 above.

103 Police (Complaints) (General) Regulations 1985, SI 1985 No 520; Police (Complaints) (Informal Resolution) Regulations 1985, SI 1985 No 671; Police (Dispensation from requirement to Investigate Complaints) Regulations 1985, SI 1985 No 672; Police (Complaints) (Mandatory Referrals) Regulations 1985, SI 1985 No 6730.

104 n 94 above, s 2.

105 *Ibid* s 3 para 3.8.

106 Disagreement between police and representatives for complainants in the Home Office Working Group on disclosure of investigation officers reports, set up as part of the Home Secretary's *Stephen Lawrence Inquiry: Action Plan*, caused the Home Secretary to delay reform in this area and declare the issue would be revisited in the wider context of independent investigation of complaints: Home Secretary, *Stephen Lawrence Inquiry: Action Plan First Annual Report on Progress* (2000) available on <www.homeoffice.gov.uk>.

to change in this area. The likely consequence in the short term is that responsibility for reforming the police complaints process is going to rest with the victims of police wrongdoing who will have to resort to law. Pressure is likely to be brought to bear laterally through actions for damages, which will be the most frequently used means of redress available, or directly by challenge to complaints, discipline or criminal prosecution decisions. In both these areas of police litigation it is probable that incorporation of the ECHR will have a significant influence on liability for police wrongdoing.

An impression of European human rights concerns with police accountability can be ascertained from the report of the CPT's delegation to the UK in September 1997, shortly before the HAC commenced its oral evidence hearings. The CPT's terms of reference were, *inter alia*, 'to carry out a detailed examination of the efficacy [sic] of existing legal remedies in cases involving allegations of ill-treatment by police officers in the United Kingdom.'¹⁰⁷

The CPT delegation took an in-depth look at the statutory police complaint and discipline processes, the use of police actions as an alternative remedy to misconduct and the judicial review proceedings of July 1997. In its report the CPT gave several grounds for doubting the efficacy of the process for prosecuting police officers.¹⁰⁸ First, the police retain a firm grip on the handling of complaints throughout the statutory process.¹⁰⁹ Secondly, the police have a substantial degree of influence over whether criminal proceedings are brought against officers who have been the subject of a complaint.¹¹⁰ Thirdly, evidence to the delegation suggested that the CPS applies an 'unduly high threshold' when considering whether to prosecute police officers.¹¹¹ Fourthly, and in summation:

in the light of all of the information gathered before, during and after the visit, the CPT has formed the view that, as matters stand, many victims of police misconduct may have little realistic prospect of other than pecuniary redress. From the standpoint of the prevention of ill-treatment of detained persons by police officers, such a situation cannot be considered satisfactory.¹¹²

In recent years, the ECtHR has interpreted the ECHR in a manner which emphasises its character as a living instrument which is capable of providing effective protection for human rights by imposing positive obligations on contracting states.¹¹³ The CPT's reference to pecuniary redress as an insufficient remedy, and the connection made which the requirement to prevent the ill-treatment of persons held in police custody, alludes to a raft of ECHR case law. In a number of cases, primarily involving Turkey and the administration of justice in Kurdistan, and including Bulgaria and France, the authorities responsible for criminal investigation and prosecution have been found wanting for their failure to take appropriate action after allegations have been made against law enforcement

¹⁰⁷ n 4 above, para 4. Under Article 8 (2)(d) of the European Convention for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment the CPT had unrestricted access to CPS, PCA and MPS files, met with senior representatives of government and criminal justice agencies and visited four London police stations (*ibid* para 5)

¹⁰⁸ Concerns were also expressed with the discipline process, which at the time of the delegation's visit were under PACE, Part IX.

¹⁰⁹ n 4 above, para 48.

¹¹⁰ *ibid* para 49.

¹¹¹ *ibid* para 50. The CPT requested the government's comments on this observation, *ibid* para 57.

¹¹² *ibid* para 53. The CPT delegation made a series of recommendations to the Home Secretary, some of them reiterating the HAC's proposals. The Home Secretary's response to the CPT Report was made available on the CPT homepage (n 4 above) on 11 May 2000.

¹¹³ On positive obligations under the ECHR see K. Stammer, *European Human Rights Law* (London: Legal Action Group, 1999) ch 5.

officers. Moreover, the ECtHR recently applied its statement on the obligation of law enforcement officials to prevent crime in *Osman v UK*, 'to do all that could be reasonably expected of them to avoid a real and immediate risk to life of which they have or ought to have knowledge',¹¹⁴ to a case involving the Turkish security forces.¹¹⁵

In these three areas – crime prevention, criminal investigation and the prosecution process – the Human Rights Act 1998 is likely to have a significant influence on police misconduct. It already has had an impact, with the Lord Chief Justice ruling that there is an expectation that the DPP will give reasons for a decision not to prosecute after a coroner's court has returned an unlawful killing verdict, unless there are compelling grounds against doing so.¹¹⁶

Until recently the state's positive obligations to protect Convention rights were largely restricted to maintenance of the article 8–11 fundamental freedoms.¹¹⁷ ECtHR jurisprudence developed significantly in this area to include the prosecution process in *X and Y v Netherlands*.¹¹⁸ A lacuna in Dutch law prevented a father from making a criminal complaint following a sexual assault against his mentally disabled 16-year-old daughter. In response to their application to the ECtHR, the Dutch government contended that the availability of civil proceedings as a means of redress ensured its compliance with Article 8. Finding against the Netherlands for its failure to institute a proper system of law capable of prohibiting crimes which violate Convention rights, the ECtHR explained:

although the object of Article 8 is essentially that of protecting the individual against arbitrary interference by the public authorities, it does not merely compel the State to abstain from such interference: in addition to this primarily negative undertaking, there may be positive obligations inherent in an effective respect for private or family life. These obligations may involve the adoption of measures designed to secure respect for private life even in the sphere of the relations of individuals between themselves.¹¹⁹

With regard to private law remedies, the Court continued:

the protection afforded by the civil law in the case of wrongdoing of the kind inflicted on Miss Y is insufficient. This is a case where fundamental values and essential aspects of private life are at stake. Effective deterrence is indispensable in this area and it can be achieved only by criminal law provisions.¹²⁰

X and Y v Netherlands has been relied on in cases involving law enforcement officers as authority that an effective prosecution machinery is in place, and developed to allow for scrutiny of decisions not to prosecute. *Aksoy v Turkey*¹²¹ is a pivotal case which drew together developments regarding injuries sustained by persons detained in custody¹²² and on the need for careful scrutiny of alleged violations by agents of the state.¹²³ The applicant¹²⁴ claimed that when in police custody he was blindfolded, stripped naked, had his hands tied behind his back and was strung up by his arms for approximately 35 minutes during which time

114 *Osman v UK* [1999] 1 FLR 198 para 116.

115 *Mahmut Kaya v Turkey* (2000) unreported.

116 *R v DPP ex p Manning* (2000) *Times* 19 May.

117 *Marckx v Belgium* (1979) 2 EHRR 14; *Plattform Ärzte für das Leben v Austria* (1991) 13 EHRR 204.

118 (1985) 8 EHRR 235.

119 *ibid* para 23.

120 *ibid* para 27.

121 (1996) 23 EHRR 553.

122 *Tomasi v France* (1992) 15 EHRR 1, *Ribitsch v Austria* (1995) 21 EHRR 573.

123 *McCann v UK* (1995) 21 EHRR 97.

124 Mr Zeki Aksoy was shot and killed in 1994 the year after he applied to the ECtHR and his father proceeded with the application, *Aksoy v Turkey* n 121 above, para 7.

electrodes were connected to his genitals and water thrown over him.¹²⁵ In *Tomasi v France*¹²⁶ and *Ribitsch v Austria*¹²⁷ violations of Article 3, inhuman or degrading treatment, were found to have occurred in police custody. In both cases it was difficult to verify allegations of assault due to the absence of independent witnesses. Under these circumstances the ECtHR determined that there had been breaches of Article 3 on account that the alleged injuries were consistent with the medical evidence, had not been caused prior to detention and could not have been self inflicted. In *Aksoy v Turkey* the deduction process relied on in *Tomasi* and *Ribitsch* was taken further to impose an obligation on the police:

where an individual is taken into police custody in good health but is found to be injured at the time of release, it is incumbent on the State to provide a plausible explanation as to the causing of the injury¹²⁸

It was decided that the Article 3 violation amounted to torture,¹²⁹ on account that the purpose of the treatment was to gather information, required some preparation and consisted of 'deliberate inhuman treatment causing very serious and cruel suffering'.¹³⁰

The careful scrutiny standard for examination of allegations that agents of the state violated the right to life was established in *McCann v UK* with the ruling that all the surrounding circumstances must be investigated in addition to the actions of those directly responsible for the use of lethal force.¹³¹ The same standard was applied in respect of an alleged breach of Article 3,¹³² with the ECtHR going further in *Aksoy* when finding that there had been a violation of Article 13:

where an individual has an arguable claim that he has been tortured by agents of the State, the notion of an 'effective remedy' entails, in addition to the payment of compensation where appropriate, a thorough and effective investigation capable of leading to the identification and punishment of those responsible and including effective access for the complainant to the investigatory procedure.¹³³

Here, authority in *X and Y v Netherlands* on the inadequacy of civil proceedings as a remedy for wrongdoing deserving of the criminal sanction has been enjoined with the need for careful scrutiny in *McCann v UK* to establish two additional obligations on the state. That offenders are identified and proceeded against follows logically from the assumption that the findings of an effective investigation will be acted on. Access to the investigatory procedure represents something of a departure. It is by intimately tying the non-derogability of Article 3, particularly with regard to torture and the vulnerable position of victims, and the absence of express exceptions that the ECtHR states access is implicit in the notion of an effective remedy.¹³⁴

The ECtHR ruling in *Aksoy v Turkey* has been referred to in many subsequent Turkish cases and a further 11 are listed in *Mahmut Kaya v Turkey*.¹³⁵ However,

¹²⁵ *ibid* para 14

¹²⁶ n 122 above.

¹²⁷ *ibid*.

¹²⁸ n 121 above, para 61.

¹²⁹ *ibid* para 65.

¹³⁰ *Ireland v UK* (1978) 2 EHRR 25 para 167.

¹³¹ n 123 above, para 150.

¹³² Under ECtHR jurisprudence the same general principles apply to Articles 2 and 3.

¹³³ n 121 above, para 98.

¹³⁴ *ibid*.

¹³⁵ n 115 above, para 98, and in two other violation of Article 3 judgments for failure to conduct an effective investigation in *Assenov v Bulgaria* (1998) 28 EHRR 652 and *Selmouni v France* (2000) 29 EHRR 403.

*Mahmut Kaya*¹³⁶ may prove to be a landmark case in its own right on account that, in addition to finding a violation of Article 2 for failure to conduct an effective investigation, the ECtHR also held that the Turkish government failed to meet its obligation to protect life.¹³⁷

Medical practitioner, Dr Hasan Kaya, and lawyer, Metin Can, were found shot dead after they accompanied two strangers to give medical attention to a member of the PKK who they claimed was wounded. The ECtHR supported the European Commission of Human Rights finding there was insufficient evidence to prove beyond reasonable doubt that state agents had been responsible for the murder, although it did conclude that the authorities suspected he was a PKK sympathiser.¹³⁸ The Commission also found that there was a strong suspicion, supported by some evidence, that the personal security of PKK sympathisers was at risk from members of the security forces or persons acting on their behalf.¹³⁹ Referring to the *Osman v UK* judgment that not every claimed risk to life requires that the authorities take operational preventative measures¹⁴⁰ the ECtHR determined that in the instant case 'strong inferences ... can be drawn on the facts of this case that the perpetrators of the murder were known to the authorities'.¹⁴¹ It found that as a doctor suspected of aiding and abetting, Kaya was at a real and immediate risk of falling victim to an unlawful attack, and that the authorities were aware, or ought to have been aware, of this risk.¹⁴²

Judgment that there had been a breach of Article 2 for failure of the authorities to take the necessary and reasonable steps to prevent risk to life was premised on three factors. Firstly, the ECtHR finding in two cases¹⁴³ in the same administrative region that an effective and independent procedure was not in place for investigating deaths involving members of the security forces.¹⁴⁴ Secondly, the consistent failure of the authorities in the region to investigate allegations of wrongdoing as found by the ECtHR leading to findings of violations of Articles 2 and 13 of the ECHR.¹⁴⁵ And thirdly, the ECtHR has found a breach of Article 6, right to a fair trial, in a series of cases¹⁴⁶ involving the State Security Courts presided over by a military judge.¹⁴⁷ The ECtHR summed up the situation in terms of the need for due legal process in a democratic society:

these defects undermined the effectiveness of criminal law protection in the south-east region during the period relevant to this case. It considers that this permitted or fostered a lack of accountability of members of the security forces for their actions¹⁴⁸

However apparent it may be that the extraordinary circumstances prevailing in Turkish Kurdistan do not compare with policing in England and Wales, the cases

136 Not to be confused with *Kaya v Turkey* (1999) 28 EHRR 1, where Mr Mehmet Kaya applied to the ECtHR on his own and his deceased brother's behalf.

137 A similar judgment was delivered in *Kilic v Turkey* (2000) unreported, given on the same day, 28 March. The circumstances of that case were that following the death of several colleagues a journalist who feared for his life sought protection from the local Governor before his murder.

138 n 115 above, para 74.

139 *ibid*

140 n 114 above, para 116.

141 n 114 above, para 81.

142 *ibid* para 91.

143 *Güleç v Turkey* (1999) 28 EHRR 121 and *Ogur v Turkey* (1999).

144 n 115 above, para 95.

145 *ibid* para 98.

146 *İncal v Turkey* [1998] EHRR 1571.

147 n 115 above, para 97.

148 *ibid* para 98.

briefly detailed above are relevant. In the 1970s to 1990s the UK was responsible for setting European general rights standards in cases connected with the suppression of armed republicanism prior to the onset of the peace process in Northern Ireland.¹⁴⁹ Turkey has now assumed this unpalatable burden. Moreover, the CPT's comments on the inadequacy of pecuniary remedies for police misconduct reflects the current state of ECHR case law, which asserts that the investigation and prosecution of criminal wrongdoing by law enforcement officers is integral to contracting states' obligations to protect individuals' rights.

Conclusions

Continued preoccupation with deliberation about who is to be responsible for the investigation of complaints, by the HAC in its 1997 inquiry in the first instance and subsequently by Home Secretary Jack Straw, will not resolve what has been one of the most enduring problems of modern policing. That is because the current debate on police complaints is largely restricted to considering who is to manage complaints, and how. Whereas the police interest, or more specifically that of police managers, has been broadly examined (in terms of employment practice, service morale, constitutional status and community relations), the needs of the complainant, whose primary objective is to have their grievance meaningfully addressed, have been woefully neglected. This imbalance is most clearly demonstrated by indifference, in reports or discussion papers published in the last decade, to the significance of the fact that roughly 80 per cent of complainants who have their complaints substantiated are curtly informed by letter that the officer responsible will be spoken to by a senior officer about his/her conduct. The inference is that the undisputed role of the complainant is to identify rogue police officers to managers and the complaint process serves a predominantly managerial function. This is despite the fact that the process also operates as the preliminary stage of the criminal process in cases where members of the public bring alleged criminal offences committed by police officers to the police's attention. With regard to such cases, the bona fide complainant is entitled, of right, to expect that the officer will be proceeded against in accordance with the law, as the office of constable requires. Anything less has serious implications for public confidence in the police and the credibility of the criminal process.

With hindsight, it was an inevitable consequence of the 'police misconduct package' introduced under sections 48 and 49 of the Police Act 1964 that complaints and actions for damages would emerge as alternative remedies, and that this would undermine the criminal liability of the police officer. What is surprising is that it took nearly 30 years for the legislation to impact on the statistics, as revealed by the steady increase in the number of successful claimants and the growing expense to the public purse in stark contrast to the relatively few complaints substantiated and the infrequency of criminal or disciplinary proceedings. Nor is it likely that introduction of a new independent body responsible for the investigation of complaints will, by itself, make a significant impression on recent statistical trends. It is contended that the growing popularity of actions for damages is on account that they more appropriately represent the interests of victims of police misconduct by satisfying their grievances, albeit by unsatisfactory financial *solatium*, than the complaints process. The simple transfer

¹⁴⁹ *eg Ireland v UK* and *McCann v UK*, nn 129 and 123 above, respectively.

of responsibility for the management of police complaints from the police to a non-police body is unlikely to affect the function of the process, and, as long as the civil courts, particularly after incorporation of the ECHR, are available as an alternative remedy, creation of a new body may prove to be a false dawn, in much the same way as the PCA failed to project a positive image.

It may appear from the above discussion that police corruption has been belittled, an unfortunate and unintended consequence of references to the police establishment's over-riding concern with corruption to illustrate their neglect of the accountability issues which affect policed communities. It is accepted that senior police officers should be responsible for exercising discipline, and it is to be hoped that the reformed police conduct and efficiency regulations will make for more professional and efficient policing. Similarly, it may well be the case that the current practice of requiring an officer to resign, thereby simultaneously curtailing the opportunity for offending and ridding the police service of undesirable elements, is the only pragmatic solution to proven misconduct of a criminal nature in particular circumstances. However, this cannot be contemplated as anything more than first aid, and is indicative of the need for the introduction of effective procedures which are capable of dealing with all forms of complaint – criminal, civil and work related – which satisfy the needs of complainants and the police.

After the Human Rights Act 1998, it is to be expected that the rights of complainants and police officers will come to the fore. As we become more conversant with the provisions of the ECHR, it is apparent that a balance will have to be struck between the rights of the victims/complainants and suspects/police officers. If, as is suggested here, it takes several more years before meaningful reform of the complaints process reaches the statute book, the ECHR will be available as an effective instrument for the victims of police misconduct to gain redress and require that criminal justice agencies act appropriately. It can be safely assumed that the courts will be called upon to rule on the rights of complainants who challenge police, PCA or DPP decisions, as has already happened in connection with a decision by the CPS not to prosecute following the death of a remand prisoner.¹⁵⁰ It is equally the case that police officers will rely on the ECHR in defence of their rights in similar fashion to the series of abuse of process cases brought by officers in connection with disciplinary¹⁵¹ and criminal proceedings¹⁵² in the late 1980s. These proceedings, irrespective of whether brought by complainants or police officers, will then inform the direction that reform of the complaints process will take in the not too distant future.

150 *R v DPP ex p Manning*, n 116 above.

151 *R v Chief Constable of the Merseyside Police ex p Calveley* [1986] 1 QB 424; *R v Chief Constable of the Merseyside Police ex p Merrill* [1989] 1 WLR 1077.

152 *R v Bow Street Stipendiary Magistrate ex p DPP* (1990) 91 Cr App R 283.

Multiple Wrongdoing and Offence Structure: A Plea for Consistency and Fair Labelling

Barry Mitchell*

Crimes come in all shapes and sizes, but relatively little work has been done on offence structure – Robinson's recent functional analysis is perhaps the one obvious exception. This article concentrates on incidents of multiple wrongdoing and suggests that the current substantive law is both inconsistent and confusing. Burglary, for example, is unnecessarily narrowly defined and should be expanded to include broadly similar scenarios. The law is confusing because it conflates qualitatively very different incidents under the same umbrella – serial killers, for example, commit the same crimes as those who kill multiple victims by one act. Not only does the law fly in the face of common sense but it conflicts with the principle of fair labelling – that crimes be defined to reflect their wrongfulness and severity – which seeks to fulfil some important functions in the criminal justice system.

Conspicuous by its absence from the growing body of reports and proposals by the Law Commission for the codification of the criminal law has been any detailed consideration of offence structure and composition. But this absence is not confined to the Law Commission's work, for there has been a relative dearth of discussion on the topic by reform bodies and commentators generally. A notable exception to this trend is the recent work of Robinson whose functional analysis of crime represents an attempt to simplify offences by distinguishing rules and principles by reference to the functions they fulfil and thereby to make the law more comprehensible and communicable to the general public, which is seen as a primary legal goal.¹

Clearly, through the appropriate parliamentary procedures, society can create whatever criminal offences it chooses, in a variety of structures and forms. Prima facie, there is no reason why these offences should necessarily conform to a particular format or model; they may differ enormously from one another in their size and degree of complexity. Focusing on situations in which a person commits what might be called 'multiple wrongdoing'² – ie where there are either multiple instances of (actual or threatened) harm to one interest or instances of harms to different interests – this article argues that the existing law is inconsistent and confusing, and on occasion flies in the face of the fair labelling principle. This argues that the law should recognise widely-felt distinctions between different forms of wrongdoing. Thus, crimes should be separated from one another and

* Reader in Criminal Justice, Coventry University.

1 See most recently Paul H. Robinson, *Structure and Function in Criminal Law* (Oxford: Clarendon Press, 1997). Robinson's proposals are discussed, albeit briefly, later in this article. D.A. Thomas also criticised law reform groups for failing to adopt a functional approach to the substantive law; see D.A. Thomas, 'Form and Function in Criminal Law' in P.R. Glazebrook (ed), *Reshaping the Criminal Law* (London: Stevens, 1978) 21–26.

2 I have chosen 'wrongdoing' rather than, say, 'wrongfulness' which is a cumbersome word, or 'harm' which has been regarded as unsuitable because the substantive law is concerned not only with injury or damage but also potential harm; see eg Alan Brudner, 'Agency and Welfare in the Penal Law' in S. Shute, J. Gardner and J. Horder (eds), *Action and Value in Criminal Law* (Oxford: Clarendon Press, 1992) 21–27.

labelled so as to reflect the nature and gravity of the offending.³ The principal argument is that the law could and should recognise more crimes consisting of multiple wrongdoing, whether the defendant appears to commit just one act (or omission etc) or more than one. This is preceded however, by a shorter argument that the scope of some existing crimes of multiple wrongdoing should be expanded.

Multiple wrongdoing and the structure and composition of current offences

Before looking at the structure of offences under the current law it is appropriate to briefly define the concept of 'wrongdoing' for the purposes of 'multiple wrongdoing' in this article. It refers to an act, omission or state of affairs,⁴ but that *per se* is not sufficient. It should also import an element of moral blameworthiness or culpability which is most clearly manifested in D's *mens rea* in relation to each element of his wrongful behaviour. The presence of a fault element ensures an adequate degree of voluntariness in the offender's conduct so that we can properly regard the behaviour as 'wrongful'.⁵ For example, in section 9(1)(b) of the Theft Act 1968 (burglary), the law requires *mens rea* (basic intent) as to both the trespass⁶ and the 'further offence'. The definition of robbery in section 8 *prima facie* suggests that a defendant charged with robbery need not intend to use force or put someone in fear etc, but since the force must be in order to steal it must be used or threatened deliberately rather than accidentally.⁷

Both at common law and in statutory form criminal offences vary considerably in the composition and complexity of their definition. Some are constituted by a single act (or omission) whereas others might be more accurately described as a venture in which the defendant commits several acts. Some are comparatively 'large' in that they contain more than one element of wrongful behaviour, sometimes involving threats to or infringements of very differing kinds of interests, arising out of several acts, and some involve more than one victim. Burglary contrary to section 9(1)(b) requires two wrongful acts – trespass in a building (or part), and some sort of 'further offence' (the infliction or attempted infliction of grievous bodily harm or theft or attempted theft). Indeed, the Criminal Law Revision Committee (hereafter 'CLRC') referred to it as a 'double offence'.⁸ The person whose house is 'trespassed' may not be the same person as the victim of the further offence. Robbery also consists of two wrongful acts – stealing and using or threatening to use force against someone. There are either two completed harms – the theft and the use of force against the person – or one complete harm and one

3 The principle is discussed in more detail later in the article.

4 Whilst many crimes consist of either an act or omission, some take the form of a prohibited state of affairs, which Glazebrook refers to as 'situational' crimes – being in possession of a controlled drug is an example – see Peter Glazebrook, 'Situational Liability' in Glazebrook, n 1 above, 108.

5 As opposed to 'harmful' or 'undesirable'. See Andrew Ashworth, *Principles of Criminal Law* (Oxford: Oxford University Press, 3rd ed, 1999) 110–112; and Andrew Simister, 'On the So-Called Requirement for Voluntary Action' (1998) 1 Buffalo Crim LR 403, 410–413. What makes decisions in cases such as *R v Larsonneur* (1933) 149 LT 542 objectionable is the lack of any voluntariness on the part of the accused.

6 Trespass to the building may, *prima facie*, be only a civil wrong, but the need for D to believe at least that he may be there without permission adds to the overall moral culpability.

7 This seems to be the prevailing view of robbery; see for example, J. C. Smith and B. Hogan, *Criminal Law* (London: Butterworths, 9th ed, 1999) 551.

8 Criminal Law Revision Committee, *Eighth Report: Theft and Related Offences* Cmnd 2977 (1966) para 76.

threat of harm. The victim of the theft need not also be the victim of the (actual or threatened) violence.

In addition, the scope of some offences is such that their essential nature may appear to vary according to the particular facts of the individual case. Robbery encompasses a wide range of situations in which the harm to either physical integrity or property interests, or both, may vary considerably, and may thus be seen primarily as either a crime of violence or a property offence or a combination of the two. Structurally, the only possible variation in robbery is that force may be either used or threatened (in order to steal). In contrast, the potential variations in burglary are much greater.⁹ Although the paradigm case is essentially concerned with protecting property interests – trespassory entry of a building together with the intent to steal or attempted or actual theft – numerous alternative forms combine infringements of property interests with intended/attempted/completed attacks on physical integrity or intended attacks against sexual integrity or property. The potential combination of elements that may constitute any one instance of the offence are such that one burglary may differ quite markedly from another in the overall nature of the offending.¹⁰ Unlike robbery, burglary requires no causal link between the two elements of wrongdoing.¹¹ D might enter V1's house as a trespasser and whilst there happen to grievously injure V2: the trespassory entry of a building merely provides the setting for the further offence.

This discussion also demonstrates that the basic nature of the multiple wrongdoing varies. Section 9(1)(a) burglary is fundamentally different from the section 9(1)(b) variety and robbery in that section 9(1)(a) contains one act whereas section 9(1)(b) and robbery inevitably require more than one. The multiple wrongfulness in section 9(1)(a) consists of a wrongful act plus a (separate) wrongful intent, and it is clear that the present law accepts that such an intent may assume particular significance. Thus, for example, possession of a controlled drug is distinguished from possession with intent to supply, which is a much more serious crime.

Occasionally, the law recognises aggravated forms of basic offences. Robbery is effectively an aggravated form of theft. Sometimes the offence label is in the form of 'aggravated [crime]' – for example, when D has with him a firearm, imitation firearm, weapon of offence or explosive when he commits a section 9 burglary, his offence becomes 'aggravated burglary' under section 10. Burglary becomes aggravated under section 10 because the possession of a firearm etc not only intensifies the victim's fear but it also brings the possibility that someone might be killed.¹² This provides the new dimension to D's wrongdoing which, although not necessary, puts him in a stronger position in which to commit the burglary.¹³

9 Similar crimes, in this respect, include public order offences such as riot, affray and violent disorder (because of the broad construction of 'violence' in case), and genocide which may consist of killing, physically or mentally harming, preventing births or forcibly transferring children.

10 Compare the paradigm burglary with a situation in which D enters a building with intent to rape. The former would clearly be regarded as a property offence whereas the latter might arguably be seen as a sexual crime.

11 The link in burglary is situational rather than causal.

12 See n 8 above, para 80. However, given that a 'weapon of offence' is defined so as to include anything which may be used to injure or incapacitate and thus may be something *prima facie* harmless, such as a handkerchief, there is a strong argument that in this instance the new dimension to the burglary need not be terribly serious.

13 Interestingly, it was argued at the time of the enactment of the Theft Bill that aggravated burglary was intended to cover those who set out armed with a firearm etc, but not those who set out unarmed and then made use of a firearm etc in the course of the burglary; (see J.C. Smith, 'Burglary Under the Theft Bill' [1968] Crim LR 295 and 367, 374. Professor Smith argued that had Parliament wanted to

Undeniably therefore, the law has for some time accepted multiple wrongdoing offences, and it appears that we have a concept of the essential or fundamental nature of these (and other) crimes. Thus, for example, the CLRC, when considering the relationship between the theft and the force in robbery, instinctively suggested that 'force used only to get away after committing a theft does not seem *naturally* to be regarded as robbery' (emphasis added).¹⁴ Nevertheless, it is important to recognise that our concepts of offences, of what would or would not constitute a particular crime, can change over the course of time. Not only is this obvious from the evolution of crimes such as robbery and burglary as set out in the CLRC's report, but is further demonstrated by comparing the CLRC's remark that '[w]e should not regard mere snatching of property, such as a handbag, from an unresisting owner as using force for the purpose of the definition [of robbery] ...'¹⁵ with the decision of the Court of Appeal in *R v Clouden*.¹⁶

A final point to be considered here is the identification of those whose concepts of crimes should influence the law. Undeniably, in the present political climate it would be foolish to suggest that the views of ordinary lay men and women should be ignored. Not only should we recognise the need to maintain a sufficient level of public confidence in the criminal justice system, we ought simultaneously to acknowledge that as a general principle people are more likely to respect (and obey) a set of laws with which they are in broad agreement. At the same time, that does not imply that ordinary people should be able to shape the detail of offence definitions. Indeed, constitutionally, the public appears to be given no more than a general right through the ballot box to express their opinions and wishes by electing a party to act on their behalf. Parliament is endowed with the responsibility of determining that detail. The ordinary layperson's concept of burglary is probably confined to entry of a building and stealing (or intending or attempting to do so),¹⁷ but there is no reason to suspect that anyone would want to reduce the scope of burglary from its present definition. Precisely what falls within or without the concept of a given crime, therefore, is almost certainly the proper domain of lawyers and law reformers, provided their opinions are not too far removed from those of the public.

'Transactions' or 'ventures'

It is clear from the previous section that offence structure is based not on single acts but on what might be called 'transactions' or 'ventures' which may consist of two or more acts. Burglary and robbery illustrate this. The law also reflects its transaction approach to offence structure through the principle that the *actus reus* and *mens rea* need only be sufficiently contemporaneous; D's *mens rea* need not

include both scenarios, section 10 would have substituted 'uses' for 'has with him'.) Yet the latter group of burglars are undeniably committing an aggravated form of burglary by the very use of the firearm etc at some stage in the venture. The likelihood of frightening those in the building and the possibility of causing death are present, and so it must be considered why such cases should not also be marked out separately from simple burglaries. If so, should they be regarded as sufficiently different from what the CLRC thought were 'proper' section 10 cases to require a third category of burglaries simply because there was no apparent contemplation or willingness to use violence at the outset?

14 n 8 above, para 65.

15 *ibid*.

16 [1987] Crim LR 56. Further illustration is provided looking at the definitions of burglary before and after the Theft Act 1968; see n 8 above, paras 68-74.

17 See also n 8 above, para 77.

be present throughout the whole time he commits the *actus reus*.¹⁸ Nevertheless, acceptance of this general rule can lead to potential problems where the venture becomes lengthy and factually complicated. The point will inevitably be reached when we must wonder whether so much has happened that it cannot sensibly be treated as one offence and must be reflected in two or more crimes.

It is also interesting to look briefly at the case law on consecutive and concurrent sentencing. Whilst the general rule is that where (what are currently) two or more crimes were committed during a single transaction the sentences will be concurrent, exceptions are made. For example, where D commits an offence against the person and a firearms offence consecutive sentences will be imposed essentially as a general deterrent.¹⁹ Yet concurrent sentences should be imposed if consecutive sentences would effectively punish D twice for what was really one crime. In *R v Coker*,²⁰ D attacked a woman, hitting her repeatedly across the face, tore off much of her clothing, and indecently assaulted her. The trial judge imposed two years' imprisonment for the indecent assault and three years consecutively for the section 47 assault, but the Court of Appeal made the terms concurrent because the two offences were so inextricably linked that this was effectively 'a crime' in which D had committed what the law technically regarded two offences. But where D has committed two or more quite distinct offences in a single venture the sentencing courts have been inconsistent. In *R v Jones*,²¹ for example, the Court of Appeal made terms of 12 months and six months imprisonment for driving whilst disqualified and driving with excess alcohol (respectively) concurrent since the two offences arose from the same incident of driving. Yet in *R v Wheatley*²² identical sentences for the same offences were made consecutive on the ground that D was a persistent offender – this may simply have a strategy adopted in order to impose a longer sentence on a defendant for whom there was little sympathy. Whilst it may well be that decisions in some cases really are inconsistent with one another, it is worth considering whether there are any relevant implications here. Might it not be that where the courts have made the sentences concurrent the aggregate of the wrongdoing could still be incorporated within a single offence without making it too clumsy or cumbersome, whereas the converse would have been true where the sentences were made to run consecutively?

Inadequacy of current law relating to multiple wrongdoing

Essentially, two criticisms can be levelled at the current law which render it inadequate or unsatisfactory, namely:

- (i) traditional concepts of offences such as burglary and robbery are defined too narrowly; and
- (ii) the current law insufficiently recognises crimes of multiple wrongdoing so that new offences which may consist of one or more act or omission etc should be created.

18 See, for example, *R v Thabo Meli* [1954] 1 All ER 373.

19 See, for example, *R v Faulkner* (1972) 56 Cr App R 594, endorsed in *R v French* (1982) 4 Cr App R (S) 57, and *R v McGrath* (1986) 8 Cr App R (S) 372.

20 [1984] Crim LR 184.

21 (1980) 2 Cr App R (S) 152.

22 (1983) 5 Cr App R (S) 417, followed in *R v Jordan* [1996] 1 Cr App R (S) 181.

As far as traditional concepts of crimes are concerned, the central criticism is that the current law is illogical and incomplete because it offers only partial protection of the interests it seeks to protect. The argument for creating new multiple wrongdoing crimes is based on the inconsistency and confusion caused by the existing menu of substantive offences, and on the implications of the fair labelling principle.

Fair labelling²³

Robinson is no doubt right to emphasise the need for the criminal law to communicate effectively to ordinary people the rules of acceptable and non-acceptable behaviour and the ways in which the criminal justice system deals with unacceptable behaviour.²⁴ In keeping with this, fair labelling surely seeks to represent to society as a whole the nature and gravity of offending. Although a principle rather than an absolute injunction,²⁵ fair (or representative) labelling states that crimes should be labelled and defined so as to reflect their degree of moral wrongfulness and relative gravity. Purely factual distinctions between offences can easily be recognised and labelled, but fair labelling can only make a valuable contribution if offences are delineated so as to reflect distinctions in moral wrongfulness.

The principle has an important rôle to play in the process of criminal justice. As Ashworth has pointed out, through its attempts to ensure a proportionate response to crime it assists in the law's educative and declaratory functions;²⁶ and of more practical importance, it significantly enhances the accuracy and reliability of D's criminal record by providing a safeguard against the risk that poor memories will fail to appreciate the precise nature of previous offending.²⁷ The need to recognise distinctions between different offences and degrees of wrongdoing is also significant to offenders in so far as it deters them from causing more harm because it might incur no greater condemnation.²⁸ Not only must justice be done, but fair labelling helps to ensure that it is seen to be done.²⁹

The substantive criminal law has always taken the view that a person's wrongfulness is not simply what he did or failed to do but includes material consequences and/or circumstances following and/or surrounding the act or omission. In those cases where no harm or damage actually ensued from D's conduct the law nonetheless seeks to reflect both the nature of the conduct and the possible or threatened consequences of it – for example through offences such as attempted murder or dangerous driving. The gravity of offences is primarily reflected in two ways, through the harm which was caused or threatened, and

23 See originally Andrew Ashworth, 'The elasticity of mens rea' in C. Tapper (ed), *Crime, Proof and Punishment: Essays in Memory of Sir Rupert Cross* (London: Butterworths, 1981) 45–70, 53–56. For further discussion and applications of the fair labelling principle, see Glanville Williams, 'Convictions and Fair Labelling' [1983] *Camb LJ* 85–95; Jeremy Horder, 'Rethinking Non-Fatal Offences against the Person' (1994) 14 *OJLS* 335; Steve Shute and Jeremy Horder, 'Thieving and Deceiving: What is the Difference?' (1993) 56 *MLR* 548–554; and C. M. V. Clarkson, 'Theft and Fair Labelling' (1993) 56 *MLR* 554–558.

24 n 1 above, Part IV.

25 See Ashworth, n 5 above, 92.

26 *ibid.* See also J. Gardner, 'Rationality and the Rule of Law in Offences Against the Person' [1994] *Camb LJ* 502, 504–506 and 512–520.

27 *ibid.* at 91, and n 23 above, 56.

28 Ashworth, n 5 above, 91.

29 See, for example, A. P. Simester and G. R. Sullivan, *Criminal Law: Theory and Doctrine* (Oxford: Hart Publishing, 2000) 32.

through D's personal moral culpability.³⁰ Nevertheless, distinctions between offences are not self-evident: there is still the question of how narrowly and precisely offences should be defined, how broad a category of harms should be permitted in any one crime, and whether intending a particular harm should be separated from causing it recklessly.

Taken to an extreme the law could distinguish in great detail between the various ways in which a victim is affected by the defendant's wrongdoing, but that would inevitably increase the number of crimes unduly. A better approach, it is suggested, is to seek to reflect qualitative distinctions between the ways in which victims are wronged. Thus, in relation to non-fatal offences against the person, Horder has argued very forcefully that if we are to define offences so as to reflect their moral character as well as their broad factual nature we must look at the way in which the victim's interests are infringed – how the victim is affected or disadvantaged by the wrongdoing.³¹ Losing a leg is very different from losing one's sight, and the substantive law should reflect this. At the same time, commensurability between offences – at least in a fairly broad sense – should be maintained, and this led Horder to suggest that an offence of intentionally causing serious injury, which is undoubtedly morally vacuous or sterile, should be replaced with one which expressly lists various forms of causing serious injury, including 'disables' and 'renders a person blind'. Thus, a compromise is reached in which the desire to avoid a plethora of very narrow crimes is achieved whilst simultaneously enabling the offence to expressly indicate the way in which the victim's bodily integrity was invaded.³²

A largely similar approach is already adopted in cases where there is no directly identifiable victim, such as in offences against the administration of justice, where the law concentrates on the quality of the (actual or threatened) wrong perpetrated against the legal system. Thus, distinctions are made between perverting and obstructing the course of justice, which are in turn distinguished from perjury. Returning briefly to crimes with identifiable victims, whereas in the context of personal violence it is sensible to distinguish between fatal and non-fatal, and between different types of non-fatal injury, a different approach is necessary in property crimes. Here the value of the property – which crudely corresponds to the gravity of the physical or mental injury to the person – is not *per se* significant because the nature of the impact on the victim will vary according to the victim's individual circumstances. Wealthier victims will obviously suffer less economic hardship by the loss than their less wealthy counterparts.

Clearly, fair labelling requires us to be able to identify the essential or distinctive nature and seriousness of D's wrongdoing. Offences of multiple wrongdoing operate on a different scale from those of single wrongdoing. What is important to us all is that offences such as burglary and robbery represent particular combinations of wrongdoing so that the precise details of individual elements of wrongdoing become less significant. The essence of robbery is the combination of stealing and violence; how much is stolen, how much force is used or threatened, and the method adopted, are relatively less crucial. One of the principal fundamental objections to the current practice of charging and convicting defendants who commit multiple wrongdoing during a single transaction is that

30 See Andrew von Hirsch and Nils Jareborg, 'Gauging Criminal Harm: A Living Standards Analysis' (1991) 11 OJLS 1.

31 Horder, n 23 above, 344.

32 *ibid* 345.

at best it could only be an inferior method of representing an offender's criminality. Supporters of the status quo might claim, for example, that a person's criminal record could be drafted so as to show that specific offences were committed on one occasion, but this is inadequate. Adding a note to my criminal record indicating that, for example, the ten murder convictions were for detonating a bomb in a busy shopping precinct³³ is unnecessarily clumsy and fails to distinguish different forms of criminality.

Before turning to examine the problems with the current law on multiple wrongdoing in detail, it is worth briefly commenting on Robinson's recent proposals based on his functional analysis of criminal law, if only to suggest why they should be rejected.

Robinson's functional analysis of crime

Robinson's argument is that the need for the law to communicate to the public can best be achieved by distinguishing the rules of conduct, ie what people actually do (or omit to do) which forms the basis of their criminal liability from 'rules of adjudication' which define the minimum requirements of liability for violating the law. These, in turn, are distinguished from a third set of rules which address the seriousness of the violation, to determine how much punishment is appropriate. However, it is a notable feature of Robinson's suggested codes that, with perhaps one exception, the offences he describes identify only a single element of harm or wrongful behaviour.³⁴

Multiple wrongdoing could be incorporated into Robinson's suggested codes,³⁵ but there appears to be some real doubts about the rationale for dividing the criminal code in this manner, for the net result is that we are left with a set of rules which are unduly simplistic. For example, mainstream offences such as burglary and robbery have effectively disappeared as such so that those whom we presently treat as burglars or robbers would have to be charged with separate crimes instead. The obvious objection to this is not so much that it would require us to alter our perception of those who break into dwelling-houses and steal etc, but that it prevents the criminal justice system from convicting the offender of a single offence which reflects the nature of his offending. This shortcoming is perhaps more clearly demonstrated by looking at what we currently regard as robbery. An important feature of such offending is the causal link between the use or threat of force and the theft, yet these would have to be separated if we accepted Robinson's proposal.

One of the reasons why Robinson's proposal is too radical is that he underestimates the capacity of ordinary lay-people to understand rules of law. There is surely no reasonable doubt that they readily understand the various forms which the concept of burglary might take. Ordinary people might not be aware of all the possible combinations and permutations of components which currently constitute different types of burglary, but they can readily comprehend the broad proposition that different combinations of wrongful behaviour involving trespass in

33 This scenario is discussed further in the section 'Incidents comprising one act or omission etc', see 403 below.

34 One apparent exception to this general characteristic is 'criminal sexual conduct' (see n 1 above, section 14, set out in Appendix A, 214) which refers to 'sexual contact', an expression which appears to be capable of encompassing multiple incidents of contact.

35 By adding further clauses to 'Adjustments to Base Grade' (article 31 in Part III of his Code of Adjudication), set out in Appendix B, Robinson, n 1 above, 235.

buildings make up the offence. My own recent surveys indicate that they are quite competent at distinguishing between culpable and non-culpable instances of wrongdoing, including the identification of elements of *mens rea* and/or justifications and excuses. Indeed, in many respects, the argument in this article is the very antithesis of Robinson's proposals. In short, effective communication of the rules of (acceptable and non-acceptable) conduct simply does not require the degree of simplification which Robinson proposes.

Inadequacies of current multiple wrongdoing offences

Whilst the fact that the law has been prepared to recognise multiple wrongdoing offences such as burglary and robbery is welcomed, it is arguable that it has been too restrictive in the way in which these crimes have been construed. For example, the various forms which burglary may take under section 9(1)(a) and (b) is in some respects rather odd. Of the possible ulterior intents in paragraph (a), only two – causing grievous injury and stealing – are required to be attempted or completed in paragraph (b). Thus, if D enters a building as a trespasser and damages property (or attempts to do so), or rapes any person (or attempts to do so), he is not guilty of burglary unless he had the ulterior intent (to damage property or rape any person) when he entered. Whilst rape or attempted rape is clearly a more serious wrong than burglary and arguably³⁶ thus merits separate recognition, the same cannot automatically be said of damaging property.³⁷ It appears that burglary is defined in the particular ways that it is simply because they reflect the kinds of situations which came before the courts,³⁸ but there is no apparent reason in principle why the offence should be limited to the particular combinations of wrongful behaviour set out in section 9 of the Theft Act 1968.

A strong case can surely be made out for extending section 9(1)(b) to include situations where, having entered as a trespasser, D damages property. There are clear parallels between theft and damage: property which has been stolen may be recovered in sound or poor condition or it may be lost forever. Damaged property may or may not be repairable. To the victim it does not appear to make any significant difference, either financially or sentimentally, whether the property has been stolen or damaged. Furthermore, whilst rape or attempted rape is inevitably a serious matter, it is arguable that section 9(1)(b) should be extended to incorporate it. One of the essential characteristics of burglary is that it reflects the desire for people to feel safe in their own homes, so that the *combination* of trespass in a building and rape/attempted rape is significant. The obvious counter-arguments are that the element of non-consensual intercourse is so significant that the word 'rape' should not be omitted from the offence label, and that the concept of burglary would become too large and unwieldy. Yet that in turn would seem to question the underlying nature and rationale of the offence. If committing (or attempting or intending to commit) certain former felonies whilst trespassing in a building is perceived as having a distinctive character to justify its existence, it is very likely

³⁶ An opposing view is discussed later in the article.

³⁷ When the Theft Bill was about to be enacted Professor Smith made the point that since rape was excluded from section 9(1)(b) because it is a more serious offence than burglary, logic and consistency demanded that damage to property should be included; see Smith, n 13 above.

³⁸ See originally Hale, *Historia Placitorum Coronae* (1736) ch XLVIII; and East, *Pleas of the Crown* (1803) ch XV. A similar comment was made by the Criminal Law Revision Committee (n 8 above, para 77).

to encompass a wide range of scenarios, unless there are good reasons for limiting the prohibited offences inside the building.³⁹ Ultimately, the question is whether our concept of burglary is sufficiently capacious to stretch far enough to include trespassory entry plus rape.⁴⁰ This is an issue which is discussed further in the article.

Robbery appears to be a more narrowly defined offence, being confined to the use or threat of force in order to facilitate a theft, although the nature and gravity of the threat/force and theft may vary considerably. Again, like burglary, robbery was construed as it was because the courts wanted to encapsulate a particular kind of scenario which was familiar to them in one relatively serious crime. But the suggestion made in the previous paragraph that a section 9(1)(b) burglary should be extended to include entry plus the destruction of property might apply equally to robbery. Destroying property effectively deprives a person of his property and thus represents the equivalent of theft. If D threatens or uses force against a person in order to destroy property belonging to another, he has surely committed an offence that is tantamount to what is currently regarded as robbery as defined in section 8 of the Theft Act 1968.

Once the idea is accepted that the existing definition of robbery could properly be expanded as suggested it is tempting to consider other combinations of wrongdoing which it might sensibly embrace. Should it, for example, include a combination of a threat or use of force with rape or sexual assault? The difference in gravity between the two elements may vary widely, but even where one element is markedly more serious than the other it may be argued that the combination of the two is what really matters. Traditionalists may argue that our concept of robbery is confined to the combination of using/threatening force in order to harm property interests. Moreover, we are again faced with the problem of deciding whether such a serious wrongdoing as rape should be merged within a wider notion of robbery, or whether (as is considered later in the article) its gravity should require us to treat such incidents as 'aggravated rape'.

The need for new multiple wrongfulness crimes

The principal contention of this article, however, is that the substantive criminal law's hitherto inadequate provision for incidents of multiple wrongdoing can only be properly addressed by creating new offences. Sometimes this may, in crude terms, effectively mean merging existing crimes into a single offence. It is helpful to be aware of three characteristics of multiple wrongdoing incidents. The offender may commit one or more acts or omissions – and the argument in favour of creating new multiple wrongdoing crimes is perhaps most apparent where there is just one act or omission etc. The individual elements of wrongful behaviour may vary both in their nature – they may threaten or undermine different interests – and they may vary in gravity.

39 It is interesting to note that, having accepted that the further offence (or, under the pre-1968 law, felony) should not be limited to theft, the Criminal Law Revision Committee recommended that the further offence in what is now section 9(1)(b) burglary should comprise the same as those ulteriorly intended in section 9(1)(a); see n 8 above, para 77 and clause 8 of the draft Theft Bill. Nevertheless, this article will argue that we should consider the relative gravity of the elements of wrongfulness and that the identification of the (multiple wrongfulness) crime should be based on the most serious element which, in this instance, would be rape rather than burglary.

40 Obviously, if burglary were considered to be big enough to include trespassory entry plus rape, the maximum penalty would surely have to be increased.

Incidents comprising one act or omission etc

If I go out on ten separate occasions and kill someone each time I am liable to conviction for ten separate murders or manslaughters. There is no difficulty in saying that I have perpetrated ten distinct offences on different occasions; I may be described as a serial killer. Yet if I detonate a bomb in a crowded shopping precinct killing ten people, I am again liable to conviction for ten separate murders. I have perpetrated a serious wrong simultaneously killing ten individuals but I have committed only one act. At the same time, when committing that one act I did so with *mens rea* in relation to each of my victims and thus merit conviction for all ten deaths. I might well say I have committed one crime, in which a single act resulted in multiple harm. The essence of my objection to being convicted of ten murders is simply that I did not commit ten separate wrongs. I would like to be convicted of something which more accurately depicts what I did; I perpetrated a wrong which resulted in ten deaths.⁴¹ Moreover, I should not be treated in the same way as if I was a serial killer and had gone out on ten separate occasions, each time killing by a fresh assault.⁴² Certainly, my criminal record should indicate what I have done with sufficient clarity to avoid confusing me with a serial killer.

The same should be true if, instead of detonating the bomb with the intention of killing those who happened to be caught in the blast, I planted the explosive in a building where I knew ten people whom I wanted to kill out of revenge, would be meeting. Although in this latter scenario the identity of the victims is more important to me than in the original, the objections raised in the previous paragraph to charging ten separate counts of murder are still applicable. In the original scenario I did not desire the deaths of any particular persons; my main aim – my direct intent – may be to make some form of political statement, but I believed it was virtually certain, inevitable, that shoppers would be killed. In the latter scenario I directly intended the deaths of ten people. Of course, there may be other people in the building who are killed in the explosion and if I thought that that would happen, as a side-effect of achieving my direct intent, then I should be criminally liable for their deaths.⁴³ In each of these scenarios my crime should be one of multiple murder.

There are fair labelling objections to convicting me of separate offences of murder which are perhaps better illustrated by some of the well-publicised major disasters in recent years – for example, the Herald of Free Enterprise, the Marchioness and Piper Alpha disasters. Not only are these described in the media in singular terms – ‘the disaster’ – but one of the reasons why they are so

41 Convicting me of a single crime in a case such as this would be consistent with public order offences such as riot, affray and violent disorder where several victims may suffer as the result of a single act by the defendant.

42 Interestingly, as part of a small qualitative survey of public opinion on homicides, a sample of 33 respondents were invited to comment on the following scenario; ‘In order to make some sort of political statement, a man plants a bomb in a busy shopping precinct and detonates it, killing a dozen people.’ Twenty-four respondents favoured creating a new crime for this type of situation, of ‘multiple homicide’; only four expressed any positive agreement with the current approach of charging the bomber with twelve separate murders. They objected to the substitution of a single ‘multiple homicide’ charge on the ground that that would anonymise the victims whose next of kin would be deeply offended. (Five respondents felt unable to give a direct reply.) See Barry Mitchell, ‘Further Evidence of the Relationship Between Legal and Public Opinion on the Homicide Law’ [2000] Crim LR 814.

43 If I did not kill these other people with malice aforethought but with some lesser form of *mens rea* my liability for their deaths could not amount to murder, and I would have effectively committed a mixture of murders and manslaughter. Situations of this sort are discussed later; see the reference to the *Att Gen's Reference* (No.3 of 1994), n 44 below.

newsworthy is that so much harm and damage arises in just one incident. The very fact that what might be no more than a moment's thoughtlessness can lead to so much harm and such serious harm is an important characteristic of our perception of what happened. Breaking it up into separate charges for each individual victim fundamentally undermines our understanding of the enormity of the incident, of what the defendant did.

A brief illustration might be helpful to identify a key objection to the current legal position. Compare the bomb-in-the-precinct scenario above with a situation in which a motorist drives too quickly, perhaps thereby causing an collision. The motorist may be convicted of driving whilst exceeding the speed limit and careless or dangerous driving. Although this obviously means that the one act of driving may result in convictions for more than one crime, each of those crimes represents the totality of his criminal behaviour. In contrast, each of the convictions I would receive for killing the ten shoppers by detonating my bomb is only a partial reflection of my criminality. Unlike the driving offences, no one murder conviction even approximates to an adequate record of my offending.

Where the kind of harm caused or threatened remains constant, the argument holds good even where the offender's culpability for each death varies so that he would (currently) be guilty of two different crimes. Suppose a situation arises akin to that in *Att Gen's Reference (No. 3 of 1994)*⁴⁴ where D stabs V1 not realising she is pregnant. V1 dies from the stab wound and V2 is born alive but dies *n* days later. This could be construed as murder of V1 and manslaughter of V2. Where D commits (what are currently) multiple offences of the same genus there is no difficulty in finding a suitable offence label – multiple murder, multiple manslaughter etc. But where the fault elements *vis-à-vis* the wrongful behaviour varies, so that he commits (what are currently) different offences of broadly similar species, an alternative label is needed. Thus, a mixture of what is currently murder and manslaughter might be called 'multiple homicide'. (The fact that some other word, 'homicide', is used rather than 'murder' or 'manslaughter' indicates a variation in fault elements.)

In the example used in the previous paragraph D caused two lots of harm – two deaths. It might have been suggested that his crime should be 'double homicide', which would imply that the number of harms etc should be reflected in the name of the crime. This would be unwise in that it implies there should be a new crime for every additional element of harm, and in the context of this example it would mean 'double homicide', 'triple homicide' etc. Faced with the desire to avoid a multiplicity of crimes,⁴⁵ the law might simply recognise some sort of new offence of 'multiple homicide'⁴⁶ to cover cases where two or more deaths had been caused by one unlawful activity. (Obviously, this principle is not confined to homicide and would apply across the substantive criminal law where other forms of harm are caused.) Although that would permit significant variations in the volume of harm within a single crime – theoretically, from two to hundreds or even thousands of deaths – the very label would clearly show that more than one person had been killed and would thus differentiate it from a single-victim homicide.⁴⁷

44 [1996] 2 WLR 412.

45 This issue is discussed more fully later in the article.

46 In my recent public opinion survey (n 42 above) respondents were invited to choose between convicting the bomber of twelve 'murders' or a single offence of 'multiple murder'.

47 Moreover, it is arguable that the concern expressed by four respondents in my public opinion survey (*ibid*) that a 'multiple homicide' offence would anonymise the victims are ill-founded, for the identity of each victim could still be made clear in the wording of the charge put on arraignment.

An offence label simply attempts to encapsulate in a very few words D's wrongful behaviour.

On the other hand, a single act (or omission) may lead to multiple harms which vary in nature. When I detonate my bomb in the shopping precinct, I may not only kill ten people but also cause extensive damage to buildings and other property. My actions were precisely the same as the earlier version which resulted in the deaths of ten shoppers but no more. The question is whether the simultaneous damage to property which, notwithstanding the high value of the property destroyed, would generally be regarded as a lesser wrongdoing, should result in a single charge rather than two charges being brought against me and, if it should, whether it had a sufficient impact on my criminality to justify incorporating it in the single charge. If the earlier argument is accepted, that killing ten people by detonating the bomb should lead to a single charge of some sort of multiple murder/homicide, then consistency surely requires the law to charge a single offence if I happen to simultaneously damage property.⁴⁸ The damage to property is *per se* a serious matter, bringing a significant new dimension to the incident and therefore ought not to be omitted from the record of my offending. (It will also be argued later that the additional wrongdoing should be of sufficient relative gravity to warrant the creation of a new crime.)

But what might my crime be called? 'Aggravated multiple murder' may seem vague, 'morally vacuous',⁴⁹ but it is difficult to offer an alternative label which more accurately depicts the full extent of my offending. The adjective 'multiple' is appropriate in an offence label where there are several incidents of the same kind of harm or wrongdoing. But where the nature of the wrongdoing varies it would clearly be inappropriate and a more suitable term might be 'aggravated'.⁵⁰ Using the label 'aggravated multiple murder' to encapsulate my criminality is justified because it implies that I knowingly killed a number of people and at the same time committed some other wrongdoing which significantly distinguished my offending from multiple homicides in which there was no additional harm.

The creation of 'aggravated' forms of existing crimes seems to work well where it is possible to identify the essential nature of the 'core' crime and the new dimension brought by the aggravation is less serious (though not so insignificant as to warrant wholly ignoring it) – for example, my bomb kills several people and causes relatively minor damage to property. The situation is more difficult where there is no dominant element of wrongdoing.⁵¹ For example, when I detonate my bomb I may kill nobody, but injure various people and damage property such that the non-fatal injuries and the damage to property are of similar gravity. Again, consistency requires us to say that if one charge is merited in the earlier scenarios where I either kill several people or kill people and damage property, then one charge should also be merited if I happen to cause non-fatal injuries instead. On that basis, two possibilities arise in evaluating the essential nature of my wrongdoing. One is to place the emphasis on what I did, ie my conduct, rather than the consequences of it, and charge me with 'detonating a bomb (perhaps in a public place)' or 'causing an explosion (in a public place)'. The idea that offences should

48 The simultaneous damage to property may be fortuitous but I may well still have envisaged it as a possible outcome of detonating the bomb and I am therefore morally culpable for it.

49 Horder, n 23 above, 339.

50 The use of labels in the form 'aggravated [offence]' would also be appropriate where D commits one or more acts.

51 As acknowledged at the outset of the article, there is, of course, always the possibility that society might recognise a particular combination of wrongdoing as crime 'X'.

be defined in a manner which ignores the consequences of D's acts is clearly not new. For example, the CLRC argued that if death is merely fortuitous motorists who kill should be prosecuted for an offence which accurately reflects the manner of their driving, rather than a 'causing death by ...' offence.⁵² But the North Committee⁵³ and Parliament disagreed, partly because of a perceived expectation that the public would want the law to reflect the loss of life.⁵⁴ Thus, the alternative approach would focus instead on the harm I have brought about and charge me with 'causing multiple injury and damage'. This is *prima facie* rather cumbersome, but it does offer a reasonably clear reflection of the range of my wrongdoing, and elsewhere the criminal law has not been averse to adopting lengthy offence labels.⁵⁵

Multiple wrongdoing based on more than one act or omission etc

Suppose that instead of possessing a firearm etc, a burglar, having entered the building, then stole not one but several items of property. Is this a different kind of 'aggravated burglary' from that envisaged in section 10 of the Theft Act 1968 because several thefts have been committed? Should the additional further offences in the building be treated as equivalent to possessing a firearm etc? Under the current law, he is theoretically liable to conviction for a s 9(1)(b) burglary and various thefts. But it is submitted that the concept of burglary should be sufficiently capacious to include multiple incidents of the further offence. The commission of several thefts in the building obviously does not necessarily imply greater harm than in cases where only one theft is perpetrated. Even if the offender stole property and inflicted grievously injury whilst in the building, he should still only be guilty of burglary because the concept of burglary should be big and flexible enough to encompass more than one instance of the further offence and (within limits)⁵⁶ even more than one kind of wrongdoing. Burglary is about intending to commit, attempting to commit or actually committing certain wrongs whilst trespassing in a building, and this would surely be a case in which the court could properly reflect the extent of the wrongdoing whilst in the building by increasing the sentence. If, on the other hand, the burglar has a firearm, imitation firearm, weapon of offence or explosive with him, that makes it aggravated burglary because the possession of the firearm etc brings a wholly new element of (threatened) harm – the possibility of killing – to the situation.

However, suppose D entered a house as a trespasser, intending to steal property and whilst there raped someone. Existing definitions of crimes such as burglary and robbery, which may both include stealing, clearly show that offences overlap with one another – sometimes they trespass on one another's territory. Earlier reference was made to the possibility that our concept of burglary might be sufficiently elastic to encompass trespassory entry of a building plus rape, but the element of rape obviously represents the most serious aspect of D's criminality.

52 Criminal Law Revision Committee, *Fourteenth Report on Offences against the Person* Cmd 7844 (1980) para 142.

53 Department of Transport and Home Office, *Road Traffic Law Review, Report* (London: HMSO, 1988) paras 6.5, 6.6, 6.9 and 6.27.

54 I might, for example, be charged with causing an explosion likely to endanger life or property (Explosive Substances Act 1883, s 2), but that clearly fails to reflect that fact that I have not just endangered life, I have extinguished many lives.

55 See eg the offence of 'going equipped for burglary, theft or cheat' defined in the Theft Act 1968, s 25.

56 It will shortly be argued that the gravity of D's wrongdoing in the house might have to be recognised in a different crime other than burglary.

Thus, if our concept of burglary is not to be expanded considerably, the focus of the criminal justice system's attention should be on the rape. When determining what constitutes the core or fundamental nature of D's wrongdoing – as a sexual offence, property crime etc – the law will concentrate on the most serious wrongdoing. This scenario would be treated as a rape which occurred in the course of a burglary – the latter bringing a new and significant dimension to the rape – so that D's crime should be one of 'aggravated rape'.

The requirements of a 'multiple ...' or 'aggravated offence'

Courts frequently deal with multiple wrongdoing situations by treating what might be called the lesser or peripheral wrongdoing as aggravating factors which simply increase the sentence for the central crime. But it is arguable that in some of these instances it would be preferable to charge a separate, 'aggravated' form of the crime. For example, the leading authority for sentencing in rape is *R v Billam*⁵⁷ where Lord Lane CJ identified several factors⁵⁸ which would normally increase the punishment, including (a) carrying on a campaign of rape; (b) perverted or psychopathic tendencies or gross personality disorder which render the offender a danger to women indefinitely; (c) the use of violence beyond what was necessary to perpetrate the rape; (d) using a weapon to frighten or wound; (e) repeated rape; (f) careful planning; (g) previous convictions for serious violence or sexual offences; (h) subjecting the victim to further sexual indignities or perversions; (i) where the victim is very young or very old; or (j) the physical or mental effect on the victim is especially serious. Factors (b), (f), (g), (i) and (j) should do no more than aggravate the sentence, for whilst they may imply greater harm or culpability compared to a 'non-aggravated' rape, none of them indicate that anything other than sexual intercourse knowingly or recklessly without the victim's consent took place. Factors (a) and (e) refer to multiple incidents of rape which may be committed on separate occasions over a period of time. On the other hand, factors (c), (d) and (h) clearly imply there was some additional harm or threat of or attempt to harm. The nature and quality of these factors take the crime beyond the basic definition of rape. They bring a significant new dimension to the incident and the offence should be one of 'aggravated rape'.

If the elements of multiple wrongdoing are of broadly similar gravity it is difficult, in the absence of society happening to recognise a particular combination of elements as constituting one offence, to see how a single definition can justifiably ignore any individual element. Each makes an equal contribution to the overall criminality. But where the elements vary in gravity, as in the previous paragraph, the situation is somewhat different. Anything which is *de minimis* can, presumably, be properly ignored, but beyond that it is difficult to offer precise arguments with any great confidence. Indeed, it was earlier acknowledged that the definition of aggravated burglary, especially the concept of 'weapon of offence', is such that the aggravation may be relatively minor so that it adds nothing of any real substance to D's burglary. But that is a criticism of that particular definition in section 10 of the Theft Act, not a criticism of the principle being argued here.

Simple logic suggests that if the lesser wrongdoing is comparatively minor it should do no more than aggravate the sentence for the dominant wrongdoing, rather than justify the creation of a separate multiple wrongdoing offence. The gap

⁵⁷ (1986) 8 Cr App R (S) 48.

⁵⁸ *ibid* 50, 51.

between the dominant and lesser wrongdoing will need to be relatively small if the lesser wrongdoing is to make a sufficient contribution to the overall wrongdoing of the incident. Suppose V1 and V2 suffer serious injuries – V1 is blinded and V2 is permanently disfigured – whilst V3 suffers minor lacerations and V4 is mildly concussed. When choosing an appropriate offence label, the lesser injuries to V3 and V4 are relatively insignificant compared to the blindness and disfigurement, and might safely be ignored. Even so, the wrongs to V1 and V2 are qualitatively different, and although it has been argued that there should be greater specificity and finer distinctions in non-fatal offences against the person, it has been acknowledged that certain circumstances may justify the use of more vague labels such as ‘wounding’ or ‘injuring’.⁵⁹ Thus, the four-victim scenario here could be regarded as ‘multiple serious injury/harm’.

Where the lesser wrong is itself a crime we might conclude that a separate multiple wrongdoing offence should be recognised where the lesser wrong falls within the same procedural classification – ie for mode of trial purposes – or within the next classification. Thus, if the dominant wrong constitutes an indictable-only crime then where the lesser wrong is indictable-only or triable either way, a multiple wrongdoing offence would be appropriate. But if the lesser wrong is merely a summary offence it should merely increase the sentence for the dominant wrong. It is arguable that this would go some way to ensuring that the gap between the two wrongs is not so great that the lesser wrong would be treated as comparatively insignificant. Of course, there can be no objectively correct means of determining when and where to create a separate crime. A situation might arise, for example, in which there is a dominant wrongdoing at the upper end of, say, indictable offences accompanied by a lesser wrongdoing at the lower end of triable-either-way offences, but since the latter comes within the next offence classification a separate multiple wrongdoing crime is probably justified.

There are though numerous situations in which the lesser wrong may not *per se* constitute a crime and there is no immediately apparent method of measuring or quantifying the two wrongs so as to automatically determine the appropriate offence structure. It was argued at the start of this article that behaviour should only be treated as ‘wrongful’ if there is sufficient fault in relation to it, so that if D lacked blame *vis-à-vis* the lesser wrong it could not be said to have added to his moral culpability and thus ought not to increase his criminality. Suppose V2 hears the impact of a collision of two vehicles and suffers mental distress in consequence. The collision, which results in the death of V1, is caused by D’s dangerous driving. If D bore no blameworthiness for V2’s injuries (because they were not reasonably foreseeable), then although there is harm in addition to V1’s death D’s crime should still be one of simply causing death by dangerous driving and V2’s injuries may be reflected in the sentence. This very point has been made in questioning the rationale behind constructive manslaughter where D has no *mens rea vis-à-vis* causing death.⁶⁰

Situations such as this, where wrongdoing of very different degrees of gravity is caused, may be seen as illustrating an interesting characteristic of offence delineation. At the comparatively lower end of the gravity spectrum, the harm gap between two crimes – for example, causing injury and causing serious injury, or

59 Horder, n 23 above, 347. Horder offered the scenario in which ‘D breaks V’s ribs, slashes V’s back with a knife, knocks V out with a heavy blow, or (without V’s consent) removes one of V’s back teeth’.

60 See *Legislating the Criminal Code: Involuntary Manslaughter* Law Com No 237 (1996) para 3.6.

causing serious injury and killing – may be quite small. But further along the spectrum, say where multiple deaths are caused, the additional presence of non-fatal injuries would be regarded as relatively insignificant to the overall scale of the incident and would surely be ignored for the purposes of identifying the essential nature and magnitude of the criminality. In other words, as we move towards the upper end of the gravity spectrum, so the seriousness of any additional harm should increase before the creation of a separate offence can be justified. But once there is sufficient additional, relatively serious harm, a separate offence *should* be recognised.

More than two different species of wrongful behaviour in one venture

It has been argued that in principle provided it is possible to identify the essential nature of the criminality, incidents of multiple wrongdoing can and should be treated as crimes in their own right, carrying their own label. However, once the facts become more complicated, especially where there are more than two different kinds of wrongdoing, labelling the incident in one offence is likely to become much more difficult, though it is possible that a particular combination of varied wrongdoing will be recognised as a crime. Suppose D enters a flat in an apartment block with intent to steal, contrary to section 9(1)(a) of the Theft Act 1968, and whilst there he rapes a woman and sets fire to her husband causing him life-threatening injuries. The fire quickly spreads to other flats in the block causing extensive damage. The rape and assault are clearly both very serious harms, more so than the burglary element, but there is also the significant destruction of property. Surely so much harm and wrongdoing of various forms has occurred in this scenario that it is impossible to reflect the essential nature – the ‘moral nominalism’ – of D’s criminality without making the label too clumsy – unless, of course, this is seen as a new variety of aggravated rape. But then suppose further that on his way out of the house D recklessly knocked over a paraffin heater setting the building ablaze. The situation becomes so factually complicated that it is impossible to reflect the criminality in a single label.

More thoughts about multiple wrongdoing offence labels

Currently, the labels by which crimes are commonly known vary considerably in that some consist of just a single word – ‘theft’, ‘rape’, ‘robbery’ etc. – whereas others are much more cumbersome – such as ‘being concerned in the management of premises for an unlawful purpose’. Whatever form they take, the aim is that the label should encapsulate and indicate the essential nature of the wrongdoing. One of the central thrusts of this article has been that common sense, consistency, the desire to avoid confusion, and fair labelling imply that new offences of multiple wrongdoing should be created. As indicated earlier, there is always the possibility that society will recognise a particular combination of wrongdoing as constituting a crime and attach a particular label to it, whether that crime arises out of one or more acts or omissions etc, or whether it consists of various kinds of (actual or threatened) harm. The label may be in a shorthand form, as where one word encapsulates the overall wrongdoing – ‘robbery’, for example, is clearly preferable to ‘violent theft’ – or if no shorthand term is identifiable the label may be a little clumsier and include an adjective such as ‘aggravated’ or ‘multiple’ to imply the

presence of more than a single element of wrongdoing. Where the wrongdoing comprises instances of the same kind of harm the crime might be labelled 'multiple [offence]'.⁶¹ Where, on the other hand, there are different kinds of harm, and especially where there is a dominant and a lesser harm, an obvious solution would be to add the prefix 'aggravated' to the existing label of the dominant wrong – thus, 'aggravated murder', 'aggravated rape' etc. Occasionally, as in the scenario where the bomber kills several people and causes considerable damage to property, there is a strong case for an offence of 'aggravated multiple [homicide]'.

The adjective 'aggravated' is *prima facie* ambiguous; it does not communicate a very clear indication of the nature of the offending, but that should not deter us from using it. What matters is that the label warns us that, for example, D's murder or rape was significantly different from other incidents of murder or rape. Only the most extreme fair labellers would want to achieve detailed levels of specificity in the definitions of criminal offences – say, between one aggravated form of rape and another. The vast majority would want to avoid the undesirable practical consequences of such extremism – the plethora of offences etc – and be satisfied that the law made a reasonable attempt to formally recognise that some rapes are significantly more serious than others because they contain much higher levels of wrongdoing.

Practical considerations

One possible objection to what has been argued here is that the increased burden of proof on the prosecution would become unjustifiable. Obviously, the greater the number of individual elements of wrongful behaviour the more work there is for the prosecution in substantiating all parts of the allegation. A defendant who had clearly perpetrated wrongdoing would escape liability if the prosecution failed to substantiate just one element in the offence definition. However, these difficulties could be avoided if certain practices were adopted. Where the multiple wrongdoing consists of what are currently separate offences, D could either be convicted of the multiple wrongdoing offence or for separate smaller crimes but not both. On the other hand, where under existing law there is only a single crime, D could be convicted of the multiple wrongdoing offence or (in accordance with current practice) for the existing offence and reflect the other wrongdoing in the sentence, but not both.

Critics might fear that the expansion of the current list of multiple wrongdoing crimes would probably cause unnecessary confusion in the minds of the jury. Yet that is surely a rather doubtful consequence of at least some of the suggestions made in this article. Expanding burglary as argued earlier does not complicate individual prosecutions any further than is likely under existing law; the offence would still consist of trespassory entry plus a further single wrongdoing (damage or destruction of property, instead of theft). Similarly, the only additional complication in offences such as 'multiple murder' is the increase in the number of victims; the basic nature of the allegation and the accused's culpability in relation to them remains constant. The prosecution would have to satisfy the jury of no more guilt than if separate counts of murder were charged.⁶² 'Multiple

61 Where there are several harms of the same kind it does not automatically follow that the crime should be called 'multiple [offence]': for example, Horder has argued that 'battery' should denote cases where D has repeatedly injured his victim; see n 23 above, 348.

62 Unless, of course, separate trials are held for each or some of the charges of murder.

homicide' is potentially more difficult in so far as it envisages different kinds of culpability *vis-à-vis* different victims. But the danger of jury confusion is only possible – and by no means likely – where there are several victims. Even then, it is doubtful that the confusion would be any greater than in many fraud trials under current law.⁶³

Fair labelling, one of the grounds relied on here for creating new multiple wrongdoing offences, is often criticised for leading to an unnecessary multiplicity of offences, each based on very narrow definitions and reflecting minute variations between factual situations. Yet that objection would not be valid in relation to what is proposed in this article. More offences would obviously be created, but offence definitions would be narrow; minute factual variations would not fall within separate definitions. Indeed, the situation would be very different. The point is that the enormous variety and nature of criminal conduct makes it inevitable that there will be a very considerable body of offences. Part of the problem, it is suggested, is that in its quest to minimise the number of offences the substantive criminal law sometimes adopts an unduly simplistic and inconsistent approach. Criminal homicide, for example, encompasses two principal crimes, murder and manslaughter, but arguably this is inadequate to reflect the full range of situations which come before the courts.⁶⁴ There appears to be some evidence of this from my recent survey⁶⁵ which indicated that on average the public favoured a greater number of offence categories in homicide than the law currently offers. This is not necessarily to imply that the law should automatically reflect public opinion in every detail, but as Robinson and Darley recently reminded us there are important reasons to ensure that the law stays broadly in step with public sentiment.⁶⁶

Conclusions

The underlying issues in this article have been concerned not so much with criminal definitions as with how we should define or conceptualise crimes. Undoubtedly, we retain some concepts of crimes, such as burglary and robbery, for very long periods, perhaps many centuries. Although the precise details may have altered slightly, we have long held the view that the worst homicides should constitute murder. At the same time, as society progresses the law creates new offences in its efforts to address what are perceived as fresh concerns.

63 In which case the preferable solution may be to try such cases before a different kind, or more appropriately qualified 'jury', as has been argued in relation to difficult fraud cases.

64 It is interesting to see that Professor Clarkson has recently proposed that the law should recognise a greater number of homicide crimes; see C.M.V. Clarkson, 'Context and Culpability in Involuntary Manslaughter: Principle or Instinct?' in Andrew Ashworth and Barry Mitchell (eds), *Rethinking English Homicide Law* (Oxford: Oxford University Press, 2000) ch 6.

65 Barry Mitchell, 'Public Perceptions of Homicide and Criminal Justice' (1998) 38(3) *Brit J Crim* 453. Undoubtedly, people found it difficult to identify distinct offence categories and there was a good deal of apparent disagreement between the categories they suggested, but what is particularly interesting here is that respondents in that survey had only been considering a limited range of homicide scenarios and yet they still favoured more offences.

66 Paul H. Robinson and John M. Darley, 'Objectivist Versus Subjectivist Views of Criminality: A Study in the Role of Social Science in Criminal Law Theory' [1998] 18 *OJLS* 409. In short, the authors argue that public sentiment is important (a) to the implementation of a criminal-penal policy based on just deserts; (b) because if people respect what the law says they are more likely to abide by it; and (c) because people tend to assume that the law broadly mirrors their moral code so that if this assumption was false they would be 'seriously confused'.

Through its adoption of a transaction approach to offence structure and its long-standing recognition of crimes such as robbery and burglary, the law has already demonstrated its acceptance in principle of multiple wrongdoing offences. Unfortunately, one of the consequences of the failure of law reformers to address the issue of multiple wrongdoing is that the current structure and composition of offences creates confusion and at times lacks common sense. This is reinforced by the fair labelling principle which implies that separate crimes should be recognised as far it is possible to describe the essential character of a criminal incident or venture and attach a label to it. What is being suggested here is the creation of a limited number of new crimes consisting of multiple wrongdoing so as to avoid confusion and provide greater consistency within the substantive law.

The essential or core identity of D's criminality is not always easy to identify in multiple wrongdoing incidents, yet superficial distinctions in the elements of wrongdoing may not necessarily make it impossible to reflect D's essential criminality in a suitable label. Only in the more complicated ventures where several different species of interests are infringed does the task become *prima facie* more difficult, but even then it is possible that society may choose to recognise a particular combination of multiple wrongdoing as an offence in its own right. Society's concept of what constitutes a crime clearly does change over time, and when it does, then fair labelling and common sense require the law to create a suitably defined crime to reflect it. Such a task ought not to be beyond the wit of law reformers and draftsmen.

LEGISLATION

The Immigration and Asylum Act 1999: A Missed Opportunity?

*Dallal Stevens**

Since 1993, the United Kingdom (UK) Parliament has passed three successive statutes on asylum and immigration: the Asylum and Immigration Appeals Act 1993, the Asylum and Immigration Act 1996, and the Immigration and Asylum Act 1999. This last, introduced by a Labour Government, is the most recent attempt to address a problem which vexed successive British governments throughout the 1990s: that of asylum. In the early months of 2000, asylum was further raised in the public consciousness as William Hague's Conservative Opposition made political mileage out of three unfortunate circumstances: first, a sudden escalation in applications with the result that, by the end of December 1999, there were 71,160 applications for that year and 102,870 claims outstanding overall, the highest numbers ever recorded;¹ second, the evident failure of the Home Office to cope with either the rising applications or the backlog; and, third, the arrival of a hijacked Afghan plane in February 2000, which, it was suggested, was part of an 'elaborate scam to claim asylum'.² The political exchanges and media coverage which accompanied these events ensured that asylum remained to the fore as an issue of growing public concern.

This review will consider the Immigration and Asylum Act 1999 and accompanying legislation in the light of recent developments and the wider political debate on asylum in the United Kingdom. Although the Act covers both general immigration law and asylum, it is the purpose of this piece not to consider the statute in its entirety but rather to concentrate on those areas which pertain to asylum law in particular.

Context

The system for dealing with asylum seekers is expensive and slow – there are many undecided cases dating back beyond 1993. We will ensure swift and fair decisions on whether someone can stay or go, control unscrupulous immigration advisors and crack down on the fraudulent use of birth certificates.³

This pledge by New Labour, in its election manifesto of 1997, was not mere political rhetoric. Decision-making was indeed slow and, as European countries struggled to cope with rising asylum applications through restrictive legislation,

* School of Law, University of Warwick.

1 Home Office, *Asylum Statistics January 2000*. These figures are provisional and exclude dependants.

2 See for example, Alan Travis, 'Straw hard line on hostages' *The Guardian* 11 February 2000.

3 *New Labour – Because Britain Deserves Better* (London: Labour Party, 1997) 35.

asylum seekers had been turning increasingly to smugglers and racketeers to gain access to European Union (EU) states. Legislative initiatives introduced by previous Conservative governments in 1993 and 1996 were largely unsuccessful in addressing the problems.⁴ Although perceived by many as unduly harsh, the statutes failed to achieve one of their avowed objectives: deterrence. From the relatively insignificant number of applications in 1987 (4,256) to the high points of 1991 and 1995 (44,840 and 43,965 respectively), asylum claims were clearly on the increase.⁵ Decisions on asylum cases had, by contrast, failed to keep pace: 2,432 in 1987; 6,075 in 1991; 27,005 in 1995.⁶ Only in 1996 did the number of decisions exceed applications (38,960 to 29,640),⁷ but by then it was arguably too late in view of the alarming backlog that had emerged (51,800 by 31 December 1997).⁸

In July 1997, the then Minister for Immigration, Mike O'Brien, announced that the newly installed Labour government would conduct a comprehensive review of immigration and asylum in the UK.⁹ Within a year, it had produced two consultation papers – the *Control of Unscrupulous Immigration Advisers* and a *Review of Appeals*¹⁰ – and a White Paper entitled *Fairer, Faster and Firmer – A Modern Approach to Immigration and Asylum*.¹¹ The White Paper introduced a wide range of proposals and envisaged the need for changes in the law.¹² A Bill, named the Immigration and Asylum Bill and largely reflecting the Paper's proposals, was eventually published on 9 February 1999. Whereas previous statutes had been relatively brief (the 1996 Act, for example, had contained only 13 sections and four Schedules), this new Bill was unexpectedly long and complex: 138 sections and 14 Schedules. In addition, the Bill still made allowance in over fifty clauses for further secondary legislation. Mike O'Brien, in addressing delegates at an Immigration Advisory Service conference in March 1999, argued that secondary legislation gave the Government a chance to 'listen'. Part of this listening process was to take place under the auspices of the Special Standing Committee established to scrutinise the Bill, a somewhat unusual step but one which had been advocated by Labour in opposition when called upon to debate the 1996 Bill.

The Special Standing Committee first met on 2 March 1999, a few days after the Second Reading, and its minutes provide a useful source of information on the major issues of concern for the Home Office and for non-governmental organisations (NGOs). Over ten weeks and 25 sittings, the Committee examined a wide range of witnesses, among them leading human rights and refugee/immigration organisations, the Local Government Association, British Airways, the Road Haulage Association, the Immigration Appellate Authority, and the Immigration Service Union, together with members of the Home Office. In addition, it received written evidence and considered each clause in detail. From a

4 See Dallal Stevens, 'The Asylum and Immigration Act 1996: Erosion of the Right to Seek Asylum' (1998) 61 MLR 207.

5 *Asylum Statistics United Kingdom 1997* (London: Government Statistical Service) Table 1.2 Note: these figures exclude dependants.

6 *Ibid.* Note: recorded decisions do not, on the whole, relate to applications made in the same time period.

7 *Ibid.*

8 *Ibid.* para 24.

9 Home Office Press Notice 166/97 10 July 1997.

10 Published respectively in January 1998 and July 1998.

11 Cm 4018 (1998).

12 *Ibid.*; see 1–4 for summary of proposals.

reading of the Committee minutes, it is clear that opposition to the Bill was both strong and widespread. Perhaps in recognition of the flaws in the Bill, the Government tabled a substantial number of additional amendments during the Committee and Report stages in both the Commons and the Lords, adding further confusion to the content of the Bill and making coherent analysis of its provisions difficult. Since some of these were quite unexpected and of far-reaching significance, such as the limitation of appeal rights in so-called 'safe third country' cases,¹³ their late introduction, and the consequent lack of Committee scrutiny, gave rise to criticism, particularly when the Government also decided to guillotine debate during the Report stage.¹⁴ Extending now to 170 sections and 16 Schedules, the Bill finally received Royal Assent on 11 November 1999, just over nine months after its First Reading. In a sense, this was just the beginning of the legislative process; the mammoth task of drawing up the numerous accompanying orders still lay ahead for the parliamentary draftsmen.

Immigration and Asylum Act 1999

The Act repeals some of the former Immigration Acts, in particular parts of the 1993 and 1996 Acts.¹⁵ Sadly, it fails to take the opportunity to consolidate and has added instead to the fragmentary nature of current immigration and asylum law. Like the first version of the Bill, the 1999 Act contains ten parts. This discussion will consider those parts which have direct implications for asylum seekers: Parts II (Carriers' Liability), III (Bail), IV (Appeals), VI (Support for Asylum Seekers), VII (Power to Arrest, Search and Enforcement), VIII (Detention Centres), IX (Registrar's Certificates: Procedures). It will also discuss relevant provisions contained in Part I of the Act and in the Schedules. The Act's provisions will be examined under select headings, rather than in the order in which they arise in the statute, beginning with a consideration of sections 11 and 12 on 'safe third countries' and proceeding to criminalisation, appeals, detention and bail, and, finally, support arrangements.

'Safe third country' cases¹⁶

Under section 2 of the Asylum and Immigration Act 1996, the Secretary of State could return asylum applicants to safe third countries on certifying that certain conditions prevailed and without substantive consideration of the claim.¹⁷ The 1996 Act also removed an in-country right of appeal in the case of EU member states or those, such as the USA, Canada, Norway and Switzerland, which had been 'designated' by order.¹⁸ Removal of the in-country appeal right reversed the principle established under the 1993 Act of in-country appeals for all. Though intended to help accelerate through the asylum process those cases deemed

¹³ ss 11, 12, 71 and 72, discussed below.

¹⁴ See for example HC Deb vol 333 col 272 15 June 1999.

¹⁵ Only ss 1, 2 and 13-16 of the 1993 Act, and ss 5, 6, 8, 12 and 13 of the 1996 Act remain. In the 1993 Act, Scheds 1 and 2 and in the 1996 Act, Sched 2 paras 1(2) and (3), 3 and 4(2), and Sched 3 paras 1, 2 and 5 are repealed.

¹⁶ Safe third country cases are those in which the applicant has travelled from the country of origin to the country in which the application is lodged, via a third country which is deemed 'safe'.

¹⁷ Asylum and Immigration Act 1996 s 2(2).

¹⁸ *Ibid* ss 2(3) and 3(2).

straightforward, it failed in this objective. Increasing numbers of applicants exercised their right to seek judicial review by challenging the Secretary of State's decision to certify, thereby delaying their departure and clogging up the appeals system.

Sections 11 and 12 of the Immigration and Asylum Act 1999 replace section 2 of the 1996 Act. Despite their far-reaching effects, they were introduced without the scrutiny of Committee (other than through the House of Lords in limited form). Section 11 of the 1999 Act allows the Home Secretary to certify claims if two quite limited conditions apply: (a) the member state to which the Home Secretary wishes to return a claimant has accepted that, 'under the standing arrangements' (ie under the Dublin Convention),¹⁹ it is the responsible state in relation to the claim; and (b) in his opinion, the claimant is not a national or citizen of the member state to which he is to be sent (s 11(2)(a)). Section 12 covers removals in all other cases. As previously, the Secretary of State has designated those countries covered by the section: Canada, Norway, Switzerland and the USA.²⁰ Certification criteria in relation to designated countries are the same as those which applied under section 2(2) of the 1996 Act.

On introducing the future section 11, the Home Secretary Jack Straw explained that 'the objective of the new clause ... [was] to deal with the problem ... of lengthy and unnecessary delays to the operation of the Dublin convention caused by disputes over the safety of transfers to other member states.'²¹ The Home Office had raised concerns about the Dublin Convention in its White Paper and had indicated therein and during Committee that it intended to improve the position.²² For some time, Home Office ministers had been complaining that entry into the Dublin Convention had been a miscalculation by the previous government. Agreed by European immigration ministers, the Convention aimed to determine the country responsible for examining an asylum claim and was intended to help resolve the unacceptable phenomenon of the 'refugee in orbit', whereby asylum seekers were shuttled to and fro between European countries as each refused to assume responsibility for processing an asylum application. When it came into force on 1 September 1997, after long delays in ratification by a number of countries, the UK found itself unable to persuade its continental partners that they should take charge of asylum seekers who had crossed the Channel.

Prior to enforcement of the Convention, the UK had been able to rely on a bilateral agreement that France would readmit asylum seekers who had simply passed through its territory. The supersession of the agreement by the Convention acted to France's advantage. When France refused to assume responsibility for a sudden influx of Roma from the Czech Republic in 1997, it was clear that the letter of the law was being applied. The Roma, despite having travelled through France to arrive in the UK, had most probably not entered the EU via France and therefore, under Article 6 of the Dublin Convention, France was entitled to refuse to co-operate with the UK authorities. In addition, under Article 11 of the Convention, the state requested to assume responsibility had to agree to do so. From the outset, it was obvious that signatories would exploit this proviso; no member state of the European Union would willingly undertake to process an

19 Convention determining the state responsible for examining applications for asylum lodged in one of the Member States of the European Communities 15 June 1990, and see below for further discussion.

20 The Asylum (Designated Third Countries) Order 2000 SI 2000 No 2245.

21 HC Deb vol 333 col 286 15 June 1999.

22 n 11 above, paras 11.24–11.30; HC Special Standing Committee Immigration and Asylum Bill col 468 22 March 1999.

asylum application unless required to do so. The UK, for the first time, found its geographical position to be a disadvantage; no longer could it rely on returning asylum seekers to its neighbours, and to France in particular, without clear proof that they had *arrived* in those countries from a non-member state (as required by Article 6 of the Convention).

Alongside problems with the Dublin Convention, Jack Straw and his Immigration Minister (now Parliamentary Under-Secretary) Mike O'Brien also faced a growing number of cases challenging the Home Secretary's certification under section 2 of the Asylum and Immigration Act 1996, and the exploitation of that provision by legal advisers through judicial review.

Asylum seekers' representatives do not find it difficult to persuade a judge to grant leave for a judicial review hearing by pointing to what they say are differences or deficiencies in the procedures in other Member States. Once leave for judicial review has been granted in such cases there can be a delay for up to a year or so as the case works its way through the courts. ... There are currently many hundreds of cases held up while the courts consider whether we can reasonably return asylum seekers to other Member States.²³

An example of such a case was *R v Secretary of State for the Home Department ex parte Lul Adan*²⁴ in which the Court of Appeal was called upon to consider three combined cases in which judicial review had been sought of the Home Secretary's decision to certify each case under section 2 without substantive consideration and to remove two of the applicants to Germany and one to France. The asylum applicants sought judicial review on the grounds that neither France nor Germany interpreted the 1951 Convention correctly, since they did not recognise that persecution from a 'non-state agent' gave rise to a well-founded fear as required under the 1951 Convention, Article 1(A)(2). This approach, known as the 'accountability theory', may be contrasted with the 'protection theory' applied by the majority of states. According to the protection theory, English courts 'recognise persecution by non-State agents ... in any case where the State is unwilling or unable to provide protection against it, ... whether or not there exist competent or effective governmental or state authorities in the country in question.'²⁵ By the time the case had reached the Court of Appeal, the Home Secretary had confirmed that he would consider the claims substantively, but the Court chose to proceed with the appeals since they raised 'a question of general importance, namely whether the Secretary of State is entitled to treat France and Germany as 'safe third countries' in relation to asylum-seekers who assert a fear of persecution by non-state agents in their country of origin, where the State is not complicit in the persecution alleged.'²⁶

In a notable decision, the Court held that the Secretary of State, in accordance with section 2 of the 1996 Act, was required to examine the practice in the third country to establish (i) if it was consistent with the 1951 Convention's true interpretation, and (ii) whether, even if consistent, it imposed such practical obstacles in the way of the claimant as to give rise to a real risk that he might be subject to *refoulement*.²⁷ The first consideration was a question of law reviewable on illegality grounds; the second was a question of fact reviewable on the grounds

²³ Letter from Mike O'Brien, Parliamentary Under Secretary of State, to ILPA 9 June 1999.

²⁴ [1999] INLR 362.

²⁵ *ibid* 377.

²⁶ *ibid* 370.

²⁷ Article 33 of the 1951 Convention prohibits the *refoulement* (ie expulsion or return) of a refugee 'to the frontiers of territories where his life or freedom would be threatened on account of his race, religion, nationality, membership of a particular social group or political opinion.'

of unreasonableness. The scope of the definition of the 'refugee' under the 1951 Convention was a matter of law. The Court judged the accountability theory applied by France and Germany to be at variance with the 1951 Convention's true meaning and the Secretary of State's certificates to have been unlawfully issued.

On the 19 December 2000, the House of Lords upheld the decision of the Court of Appeal in *Lul Adan*.²⁸ It held that there was one autonomous interpretation of Article 1(A)(2) of the 1951 Convention: namely, that the explanation provided in *Adan v Secretary of State for the Home Department*²⁹ was correct and should be applied. Sadly, Lord Steyn, who gave the leading speech, stated inaccurately that the House of Lords in *Adan* had 'authoritatively rejected the persecution theory' (that is, the protection theory) and 'adopted the accountability theory'. This is clearly wrong. As Lord Hutton rightly observed in *Lul Adan*, Lord Lloyd had confirmed in *Adan* 'that in the United Kingdom the proper construction of the Geneva Convention requires the acceptance of the persecution theory' (although *Adan* imposed some limitations to the application of that theory). This confusion aside, the House of Lords in *Lul Adan* concluded that the Secretary of State had been wrong to proceed on the dual assumption that there was a band of permissible meanings of Article 1(A)(2) of the 1951 Convention and that the practice adopted by Germany and France fell within the permissible range. The Secretary of State was held to have materially misdirected himself and his decisions were quashed.

The effect of section 11, however, is to bypass judgements such as *Lul Adan* since it no longer requires the Home Secretary to certify that the country to which the applicant is to be returned is 'safe' (though section 12 continues to do so). Rather, the Home Secretary simply has to certify that the member state has, under the 'standing arrangements', accepted responsibility for the asylum application, and that he does not believe the applicant to be a national or citizen of that state. This should, in theory, reduce substantially the number of applications for judicial review. However, in view of their significance it is unfortunate that sections 11 and 12 were not debated in Committee and received relatively little discussion.

Appeal rights in third country cases have also altered. Under section 71(2), a recipient of a certificate may appeal against it to an adjudicator 'on the ground that any of the conditions applicable to that certificate was not satisfied when issued, or has since ceased to be satisfied.'³⁰ This is very similar to the previous position under section 3(1)(a) of the Asylum and Immigration Act 1996. However, it is unlikely that many applicants will be able to make use of section 71 in Dublin Convention or designated country cases, since it can be assumed that certification will be in order. In addition, there are further limitations imposed by section 72: where an individual is to be returned to a member state or to a designated country (under section 12(1)(b)), there is no *in-country* right of appeal. In all other cases, there is such an appeal.

In the 1999 legislation, the Government made a concession to the Human Rights Act 1998. Section 65 created a new right of appeal to an adjudicator on human rights grounds. Breach of human rights relates to the action or failure of action by an authority in a way which is made unlawful by section 6(1) of the Human Rights Act 1998.³¹ Importantly, there are no limitations to section 65 and therefore a human rights appeal could be made whilst in-country. The potential use of this

28 *R v Secretary of State for the Home Department, ex parte Adan*. <<http://www.publications.parliament.uk/pa/ld200001/ldjudgmt/jd001219/adan-1.htm>>.

29 [1999] 1 AC 293.

30 Note: the appeal right is limited to an adjudicator only.

31 s 65(2) See below for further discussion.

provision in safe third country cases is evident. The Government, however, recognising the risk tried to close the loophole in section 72: in section 11 or 12(1)(b) safe third country cases,³² section 72(2)(a) provides that there is no in-country right of appeal prior to removal under section 65 where the Secretary of State has certified that the allegation that an authority acted in breach of the applicant's human rights is manifestly unfounded. From Jack Straw's words to date, it would appear that he is determined to ensure proper enforcement of the Dublin Convention and that he will therefore judge any appeals under section 65 as manifestly unfounded where they involve a member state, despite such cases as *Lul Adan*. The likely outcome of this approach is, of course, judicial review of his certification that such cases are manifestly unfounded, since Germany and France would arguably be in breach of their obligations under the European Convention of Human Rights (ECHR), and, in particular Article 3.³³ By closing one door on judicial review, the Government may have inadvertently opened another.

Increasing criminalisation

One of the priorities of the White Paper, and subsequently the 1999 Act, related to detection, prevention and punishment of immigration and asylum law offenders. Earlier statutes had instigated the use of immigration policing by non-immigration personnel; the 1999 Act followed suit by enhancing previous legislation (employers' liability) and by introducing new areas of immigration control (marriage registrars' duty). It also radically increased investigatory powers, enforcement and sanctions.

Employers' liability

Section 8 of the Asylum and Immigration Act 1996 raised concern at the time of its implementation. It created a new criminal offence for employers to employ persons who required leave to enter or remain and who were not permitted to work. All employers were forced to check the documentation of future employees to ensure that they complied with the legislation. Failure to do so could result in a maximum fine of £5,000. The accompanying regulations did, however, provide a defence: namely, that the employer had asked the employee to produce a document (as specified by order)³⁴ before commencement of the employment, and that the original document had been kept or copied by the employer.³⁵ Guidance was published by the Home Office to assist worried employers and to ensure that the immigration provision was not used as an excuse for racial discrimination.³⁶

In opposition, Jack Straw argued against section 8 and promised prior to the 1997 election that a Labour government would not enforce it,³⁷ thereby raising expectations that Labour would repeal section 8 when elected to govern. The White Paper soon dispelled that notion. There the Government made clear that section 8 would remain on the statute book but that steps would be taken 'to re-emphasise to employers their duty to avoid racial discrimination in their

³² ie involving Member States or designated countries.

³³ Article 3: 'No one shall be subjected to torture or to inhuman or degrading treatment or punishment.'

³⁴ Immigration (Restriction on Employment) Order 1996 SI 1996 No 3225.

³⁵ Asylum and Immigration Act 1996 s 8(2).

³⁶ *Prevention of Illegal Working: Guidance for Employers* (London: Home Office Communication Directorate, December 1996).

³⁷ For example Stephen Grey and Matthew Prescott, 'Straw signals U-turn to ditch tough immigration laws' *Sunday Times* 13 April 1997.

recruitment practices when seeking to secure the defence under section 8.³⁸ This they did through amendment of section 8 by section 22 of the 1999 Act. Section 22 inserts section 8A into the 1996 Act; it requires the Secretary of State to issue a code of practice 'as to the measures which an employer is to be expected to take, or not take, with a view to securing that, while avoiding commission of an offence under section 8, he also avoids unlawful discrimination' (s 8A(1)). At the time of writing, the code of practice was not available.

'Sham marriages'

Following on the theme of immigration control by non-immigration officials, the 1999 Act has introduced a rather bizarre new duty: the duty on civil marriage registrars to report 'suspicious marriages'.³⁹ While there exists a procedure whereby marriage registrars may report suspicious marriages to the Registrar General, the new provision now imposes a *statutory* duty on registrars to do so. The duty relates to all civil marriage registrars and to those conducting religious marriages, except where the marriage takes place in the Churches of England and Wales (sections 24(1) and (3)). A 'sham marriage' is defined as a marriage entered into between a person ('A') who is neither a British citizen nor a national of another European Economic Area state, and another person (whether or not such a citizen or national), and entered into by A for the purpose of avoiding the effect of one or more provisions of UK immigration law or the immigration rules (section 24(5)). Registrars will be issued with guidance on how to recognise a 'sham marriage', guidance which will, one suspects, be sorely needed.

Commentators have been surprised by this new provision. Soon after the 1997 general election, the Government ended the controversial 'primary purpose rule'. Simply put, this rule imposed an onus of proof on an applicant spouse or fiancé(e) that the primary purpose of marriage was not to gain entry to the UK. Announcing the cessation of the primary purpose rule, Jack Straw described it as 'arbitrary, ineffective and bureaucratic'.⁴⁰ It is puzzling, therefore, that he should then proceed to support the 'sham marriages' clause when it clearly echoes the primary purpose rule. The explanation provided by the White Paper was not very illuminating: 'there is ample evidence to show that large numbers of bogus marriages are being contracted in the UK every year. A bogus marriage is one arranged for the sole purpose of evading statutory immigration controls.'⁴¹

Alongside the view that section 24 is primary purpose in another guise,⁴² there is also a view that it is potentially in breach of Articles 8 (the right to respect for private and family life), Article 12 (the right to marry) and Article 14 of the ECHR (protection from discrimination on the ground of, *inter alia*, religion). The Memorandum of Evidence by the pressure group Justice to the Special Standing Committee pointed out that 'in order to justify the interference with Article 8 ECHR that the provisions provide, the Bill should specify much more precisely the circumstances which might denote a "sham marriage"'.⁴³ This suggestion was not adopted. In addition, it was felt that the exclusion of marriages carried out in the Churches of England or Wales was seriously problematic and likely to engage Article 14, since all other religious marriages requiring civil preliminaries were

38 n 11 above, para 7.12(iv).

39 s 24, and see also Part IX and Sched 14.

40 Home Office Press Notice 134/97 5 June 1997.

41 n 11 above, para 11 4.

42 See HC Special Standing Committee Immigration and Asylum Bill col 201 17 March 1999 (ILPA).

43 *ibid* col 59 16 March 1999 (Justice).

covered. The Explanatory Notes accompanying the Act state that 'The Act does not extend to marriages after ecclesiastical preliminaries because the procedure involved is considered already sufficiently detailed to ensure that the system is not abused for immigration advantage. There is no evidence that religious marriages after ecclesiastic preliminaries are abused for immigration advantage.' In the opinion of Justice, comments such as these appear to confirm rather than mitigate the discrimination. 'Any provisions that are to be applied, should be applied indiscriminately to all marriage ceremonies.'⁴⁴

Deception and facilitation of entry

The immigration offence of entering the UK illegally, contained in section 24 of the Immigration Act 1971, has become an apparent source of anxiety for governments of left and right. Both have seen fit to tamper with section 24, with the result that it now seems to cover every possibility. Section 24(1)(a) originally referred to those attempting 'knowingly' to enter the UK in breach of a deportation order or without authority. Section 4 of the Asylum and Immigration Act 1996 inserted section 24(1)(aa) which added the proviso that it was also an offence if anyone obtained or tried to obtain leave to enter or remain in the UK by any means including deception. Section 28 of the 1999 Act inserts a new section 24A into the 1971 Act and replaces section 24(1)(aa). Section 24A(1) reads:

A person who is not a British citizen is guilty of an offence if, by means which include deception by him-

- (a) he obtains or seeks to obtain leave to enter or remain in the United Kingdom; or
- (b) he secures or seeks to secure the avoidance, postponement or revocation of enforcement action against him.

The reason for the amendment is to ensure that those who seek to remain in the UK on the basis of an unfounded asylum claim involving the use of deception will be punished on summary conviction with a maximum statutory fine (£5,000), or with imprisonment for up to six months, or both. ILPA has expressed grave doubts about this new offence. Describing the section as 'quite staggering' and 'far beyond its stated aim', ILPA has 'real fears that it will criminalise asylum seekers in a way that is wholly unwarranted'.⁴⁵ There is also a possibility that the new section 24A will give rise to breach of Article 31 of the 1951 Convention in cases where an asylum seeker is prosecuted under it, only to be determined a refugee later. (Article 31 prohibits the imposition of penalties on refugees on account of their illegal entry or presence when they have come directly from a territory where their life or freedom was threatened, provided they present themselves without delay to the authorities and show good cause for their illegal entry or presence.)

For some time, media reports have focused on the burgeoning problem of the so-called 'immigration racketeer'. Faced with a tightening ring of restrictive immigration and asylum legislation in the European Union, desperate people have been turning to unscrupulous agencies to obtain assistance in gaining entry to their country of choice in the EU and northern states. It is now clear that huge financial rewards are to be gained from human smuggling and, consequently, it has become a favoured activity of various criminal groups around the world, with potentially terrible consequences – such as the 58 Chinese migrants found dead in the back of a Dutch lorry at Dover in June 2000. The Government has long sought to clamp

⁴⁴ *ibid* col 60.

⁴⁵ *ibid* col 204 17 March 1999 (ILPA).

down on immigration racketeers. In January 1998, in line with its manifesto commitment, it published a consultation document on the *Control of Unscrupulous Advisers*. In July 1998, in the White Paper, it declared its intention to take a tougher approach to deterring and preventing the arrival of inadmissible passengers,⁴⁶ and in February 1999, several sections of the Immigration and Asylum Bill were dedicated to tackling the problem of smuggling.

One of the first provisions to deal with this issue is section 29 ('facilitation of entry'), which is not new to immigration law and, in fact, simply amends section 25 of the Immigration Act 1971. Section 25 was, and continues to be, concerned with assisting illegal entry. In 1996, in order to combat the rising problem of immigration racketeering and smuggling, it was amended to include two new offences. Section 25(1) of the Immigration Act 1971 now reads that it is an offence knowingly to facilitate the entry of illegal entrants, knowingly to facilitate for gain the entry of asylum claimants, and knowingly to facilitate the obtaining of leave to remain by deception. In another bid to be 'tough on crime', section 29(2) of the 1999 Act increases the maximum custodial sentence from seven to ten years following conviction on indictment. Section 29 also adds further clarification to section 25 of the 1971 Act by exempting from prosecution 'anything done to assist any asylum claimant by a person in the course of his employment by a *bona fide* organisation, if the purposes of that organisation include assistance to persons in the position of the asylum claimant' (section 29(3)). The previous amendment undertaken by the 1996 Act had been confusing since it seemed only to exempt *bona fide* refugee organisations, thereby opening the possibility that legal firms, and those offering legal advice, might be subject to prosecution. Section 29(3) attempts to address this deficiency. Finally, it extends the definition of 'asylum claimant' beyond those applying under the 1951 Convention to those who claim breach of the ECHR.

Carriers' liability

Since 1987, and the introduction of the Immigration (Carriers' Liability) Act, the issue of carriers' liability has been a constant feature of UK asylum policy. The 1987 Act provided one of the first modern-day examples in the refugee field of the transfer of immigration policing to individuals or organisations outside the Immigration Service and Home Office.⁴⁷ It granted the Secretary of State the power to demand a specified sum from a carrier which had transported a passenger not in possession of valid documentation, whether an asylum seeker or not.⁴⁸ The original sum was set by section 1(1)(b) of the Act at £1,000 per person; in 1991, it was raised to £2,000.⁴⁹ Despite the fact that actual payments under the Act have not always been forthcoming – by 31 May 1998, for example, £117.3 million had been levied, £78.3 million had been paid, £28.7 million had been waived following representations from carriers, and £10.3 million was outstanding – it has continued to be an important pillar of Home Office policy.⁵⁰ (The discretion to waive

46 n 11 above, para 5.16.

47 Interestingly, the Aliens Act 1793 contained a provision whereby masters of vessels could be fined if they did not declare the numbers, names, and occupations of any foreigners on board their vessels: see Dallal Stevens, 'The Case of UK Asylum Law and Policy: Lessons from History?' in F. Nicholson and P. Twomey, *Current Issues of UK Asylum Law and Policy* (London: Ashgate, 1998).

48 For a wider discussion see A. Cruz, *Carriers' Liability in the Member States of the European Union* (Brussels: Churches Commission for Migrants in Europe, 1994); E. Feller, 'Carriers' Sanctions and International Law' (1989) 1 *International Journal of Refugee Law* 48.

49 Immigration (Carriers' Liability Prescribed Sum) Order 1991 SI 1991 No 1497.

50 HC Deb vol 315 WA col 427 7 July 1998.

payments was clarified in 1987 in a document entitled: *The Immigration (Carriers' Liability) Act 1987: Use of Discretion by the Immigration Service*. So, for example, under paragraph (iii), charges will be set aside where 'the passenger, who was subsequently accepted as a genuine refugee, had been unable to get the proper documents before travelling here and this country was accepted as the right and only practicable haven for him.'

As the types of carriers used to gain access to the UK have changed, so has the legislation. Thus, in 1998, the application of the 1987 Act was extended to include not only ships and aircraft but also passenger trains from Belgium by virtue of the Channel Tunnel (Carriers' Liability) Order.⁵¹ (The Order did not cover Eurostar services from Paris due to 'legal difficulties in France'.)⁵² The 1999 Act, in repealing the 1987 Act and replacing it with sections 40 to 42, added road passenger vehicles to the list of craft which could be targeted for penalties under the legislation (section 40(1)). It now reads that the Secretary of State may charge an owner £2,000 in respect of any person arriving without a valid passport or identity document, or requisite visa, on the owner's ship, aircraft, vehicle or train (sections 40(1) and (2)). In order to guarantee payment of the charge, the Act also includes a power to detain the transporter in which the undocumented person is carried (section 42). There are defences provided in subsections 40(4) and (5) which recognise, for example, that the correct documentation may have been shown by the traveller on his or her embarkation (section 40(4)). Train operators and owners or operators of road vehicles are exempted from charges if they can show that they have in place satisfactory arrangements to ensure that undocumented passengers are detected, and that reasonable steps are taken in accordance with the arrangements to check documentation and to prevent the arrival of an undocumented individual (section 40(5)). This defence applies only in circumstances where the operator/owner is unable to demand documentation from passengers due to legal restrictions in the country of embarkation (section 40(5)(a)).

Dissatisfied with these changes in relation to carriers' liability, the Government championed further, and more radical, powers. These materialised in section 32 of the 1999 Act as a new civil penalty for those responsible for the transport of clandestine entrants to the UK. This new penalty differed from that imposed by the Immigration (Carriers' Liability) Act 1987 in that it targeted *clandestine* entrants as opposed to passengers known to the transporter. A clandestine entrant is defined as someone who either arrives in the UK concealed in a vehicle, ship or aircraft, or passes or attempts to pass through immigration control concealed in a vehicle, or who arrives in the UK on a ship or aircraft, having embarked concealed in a vehicle when outside the UK; the concealed person must also claim or intend to claim asylum in the UK, or evade or attempt to evade immigration control (section 32(1)). The heart of the provision is contained in subsection 32(2), which empowers the Home Secretary to demand payment of a prescribed sum (£2,000) within a prescribed period (60 days).⁵³ Additionally, the Act contains a power to detain vehicles, small ships and aircraft in connection with penalties imposed under section 32 where a senior officer believes there to be a significant risk that the penalty will not be paid within the prescribed period (sections 36(1) and (2)).

⁵¹ SI 1998 No 1015.

⁵² n 11 above, para 5.14.

⁵³ The Carriers' Liability (Clandestine Entrants and Sale of Transporters) Regulations 2000 SI 2000 No 685 regs 3 and 4.

Though there are certain defences to the imposition of a penalty, these will be very difficult to establish.⁵⁴

The Freight Transport Association (FTA) and the Road Haulage Association (RHA), amongst others, fought hard against the introduction of these measures, and continue to object vociferously, believing them unfairly to target their members. The suggestion by Mike O'Brien, in Special Standing Committee, that the numbers of illegal entrants to the UK in the backs of lorries had increased (he cited the figure of 8,000 for 1998) cut no ice with the associations.⁵⁵ In their view, the driver and the operator were 'the first victims of [the] problem' and were unable to carry out the checks of their vehicles as required by the Government.⁵⁶ Furthermore, Steven Norris of the RHA suggested that the defences provided for by the legislation were unhelpful; once an illegal immigrant was detected, it would appear 'axiomatic' that the driver had failed to observe one of the principles contained in section 34(3).⁵⁷ Justice identified a number of further concerns. In particular, it pointed out that:

The new, and indeed the existing, penalties raise questions under the UK's responsibility to provide protection under the 1951 Convention and Article 3 ECHR. The new clandestine entry penalties are also likely to raise problems of compliance with the requirements of Protocol 1(1) ECHR in respect of interference with property rights and the Article 6 ECHR requirements for due process in determining civil obligations and criminal penalties.⁵⁸

The Home Office has now produced a Code of Practice for Vehicles which 'sets out the measures to be taken and the procedures to be followed by persons operating a system for preventing the carriage of clandestine entrants to the United Kingdom, in respect of vehicles.' These include measures to secure vehicles against unauthorised entry (para 1.1), such as sealing and/or locking a vehicle, and measures to be taken immediately prior to embarkation to the UK (para 1.2). Despite the advice, within days of the civil penalty coming into force on 3 April 2000, the Home Office had served 30 notices of penalty in respect of 30 vehicles.⁵⁹

New powers for immigration officers

A final example of increased enforcement powers is contained in Part VII of the Act. Here, immigration officers are granted unusually wide powers of entry, search and arrest with or without warrant, and fingerprinting.⁶⁰ While these enforcement powers have long been available to the police, the 1999 Act substantially increases the powers of immigration officers, making them, in effect, substitute police officers. Thus they are entitled to use reasonable force when exercising their powers under the Immigration Act 1971 or the 1999 Act (section 146), and are subject to the Police and Criminal Evidence Act 1984 ('PACE') code.⁶¹ This

54 ss 34(2) and (3); the carrier will need to show either that he (or an employee) was acting under duress, or that he did not know and had no reasonable grounds for suspecting that a clandestine entrant was concealed in the transporter, that an effective system for preventing the carriage of clandestines was in operation, and that the system was operated properly on the occasion in question.

55 HC Special Standing Committee Immigration and Asylum Bill col 409 18 March 1999.

56 *ibid* col 405.

57 *ibid* col 413.

58 *ibid* col 61 16 March 1999.

59 HC Deb vol 348 col 441 12 April 2000; the Conservative Party remains firmly opposed to this part of the Act.

60 ss 128–132; 134–135; 141; and see ss 137, 138 and 140 for details on safeguards for search warrants (similar to the Police and Criminal Evidence Act 1984 (PACE)), the execution of warrants (again similar to PACE), and the extension of power to grant an immigration officer as well as a constable a warrant to search premises and arrest someone liable to detention.

61 On 7 February 2000, the Immigration (PACE Codes of Practice) Direction 2000 came into force.

massive increase in powers rang alarm bells for some non-governmental organisations. One query related to the extent of training to be received by the Immigration Service. In response, it was suggested by the Government in Special Standing Committee that a select group of about only 20 immigration officers would be trained to use the wider powers.⁶² None the less, members of the Public and Commercial Services Union (the union for the Immigration Service) balked at the idea of such extensive powers, despite the promise of police training. The secretary to the immigration staff branch of the Union was reported as saying: 'The immigration service is largely a bureaucratic organisation, but this will lead to a dramatic change of emphasis to its role. We are worried about the safety of immigration officers. What happens if things go wrong after arrest? This is not what we are about. We fear race relations could suffer.'⁶³ Irrespective of such wider implications, this sets a worrying precedent whereby the burden and cost of policing is shifted on to another public body.

Article 31(1) of the 1951 Convention

In contrast to the numerous parts of the 1999 Act providing for increased sanctions and enforcement, section 31 introduced a pro-refugee clause very late in the Parliamentary proceedings. It creates a defence against charges for certain offences (set out in sections 31(3) and (4)) where the refugee, having come to the UK directly from the country where his or her life or freedom was threatened (within the meaning of the 1951 Convention), presents himself or herself to the authorities in the UK without delay, shows good cause for his illegal entry or presence, and makes a claim for asylum as soon as is reasonably practicable after arrival (section 31(1)). This defence mirrors Article 31(1) of the 1951 Convention. The requirement that the refugee should come directly from the country in which he or she is threatened is further explained by section 31(2), which makes it clear that if an individual stops in a country outside the UK in which he or she could reasonably have been expected to be granted asylum, the defence contained in section 31(1) does not apply. It should also be noted that, like the 1951 Convention, section 31 refers to 'refugees' and not to asylum seekers.

The late entry of this provision to the Bill was due to judicial intervention. In July 1999, the Queen's Bench Division handed down a highly significant judgement in *R v Uxbridge Magistrates' Court and Another ex parte Adimi*.⁶⁴ In this case, the three applicants, Adimi, Sorani, and Kazuu, all asylum seekers, had been charged with offences relating to the possession of false passports and other documents used to travel to the UK. Simon Brown LJ opened with a stern criticism of current Home Office policy: 'First ... I pause to note one striking fact. The respondents acknowledge that, until these challenges were brought, no arm of State, neither the Secretary of State, the Director of Public Prosecutions, nor anyone else, had apparently given the least thought to the UK's obligations arising under Art 31.'⁶⁵ 'What, then, was the broad purpose sought to be achieved by Art 31?' he queried.

Self-evidently it was to provide immunity for genuine refugees whose quest for asylum reasonably involved them in breaching the law. ... That Art 31 extends not merely to those

⁶² HC Special Standing Committee Immigration and Asylum Bill cols 1474–1475 13 May 1999.

⁶³ Marie Woolf, 'Immigration officers balk at issue of handcuffs and CS spray' *Independent on Sunday* 10 May 1999.

⁶⁴ [1999] INLR 490.

⁶⁵ *ibid* 495

ultimately accorded refugee status but also to those claiming asylum in good faith (presumptive refugees) is not in doubt. Nor is it disputed that Art 31's protection can apply equally to those using false documents as to those (characteristically the refugees of earlier times) who enter a country clandestinely.⁶⁶

Simon Brown LJ then went on to consider the three caveats to the application of Article 31: 'coming directly'; 'present without delay'; 'good cause'. He found that all three applicants should have been exempt from penalty, though he declined any form of declaratory relief. 'Rather,' he concluded, 'I would let our judgments speak for themselves, with a view to ensuring that the mistakes of the past are not repeated in future. It must be hoped that these challenges will mark a turning point in the Crown's approach to the prosecution of refugees for travelling on false passports. Article 31 must henceforth be honoured.'⁶⁷ The very evident admonition contained in this judgement was acknowledged by the Government when it introduced section 31 during the Report Stage in the House of Lords.

Appeals process

Appeals under the Asylum and Immigration Appeals Act 1993 and the Asylum and Immigration Act 1996

It is safe to say that the one area which has been tampered with more than any other in the 1990s is that of appeals. The Asylum and Immigration Appeals Act 1993 compensated for the long-time lack of an asylum-specific appeal.⁶⁸ It provided a right of appeal to all asylum seekers for the first time, and opened up access to the higher courts, the Court of Appeal and House of Lords. The new system introduced in 1993 provided a two-tier appeals system: the 'fast' and the 'standard' track. The fast track was concerned only with so-called 'without foundation' cases, that is cases which were deemed to be 'frivolous or vexatious', or 'which did not call into question the UK's obligations under the 1951 Convention'. Strict time limits were imposed on without foundation cases, and there was no further right of appeal beyond the first hearing with a special adjudicator. Following the 1993 Act, most without foundation cases were found to be those involving 'safe third countries'.

Though the success of the 1993 Act was questionable,⁶⁹ the drafters of the Asylum and Immigration Act 1996 were keen to extend the range of cases subject to the fast-track procedures. Under section 1 of the 1996 Act, a number of new categories were added, including the infamous 'White List'. The White List comprised a list of countries designated by the Home Secretary as presenting no serious risk of persecution. The general effect of the amendments was to ensure that most asylum claims would be dealt with by the accelerated appeals process, and would have only one appeal to a special adjudicator. The time limits continued to be very tight: between two and seven days, depending on the circumstances.⁷⁰ An exemption was provided where there was a reasonable likelihood that the appellant had been tortured in the country or territory to which he or she was to be

⁶⁶ *ibid* 496.

⁶⁷ *ibid* 506.

⁶⁸ Pre-1993 appeal rights were dependent upon the status of the asylum applicant at the time the application was lodged, and there was no appeal to the Court of Appeal and House of Lords.

⁶⁹ See for example report by KPMG Peat Marwick, *Review of Asylum Appeals Procedure* (London: Home Office/Lord Chancellor's Department, 15 December 1994); The Glidewell Panel, *The Report from an independent enquiry into the implications and effects of The Asylum and Immigration Bill 1995 and related social security measures* (London, 16 April 1996).

⁷⁰ Asylum Appeals (Procedure) Rules 1996 r 5. See for further discussion n 4 above.

sent.⁷¹ The 1996 Act also removed the in-country right of appeal for certain safe third country cases (involving EU member states and 'designated' states), thereby withdrawing a guarantee which had been granted under the 1993 Act.

The White List created a stir, not only among non-governmental organisations but also among MPs of the then Labour Opposition. Promises to remove the White List if elected to government were included in the White Paper,⁷² but the Government went on to add that, since it felt that no unfairness had resulted from the White List in practice, it would continue in operation until new appeals arrangements were implemented.⁷³ This was disappointing for those favouring immediate abolition. The Government was nonetheless intent on implementing radical changes to the appeals process, and published a consultation paper in July 1998 entitled *Review of Appeals*. Its main task was stated as being 'to investigate whether the current appeal system provides adequate rights of appeal consistent with the costs to public funds and its impact upon the operation of immigration control'.⁷⁴ The review claimed that: applicants tended to exercise multiple appeals to delay removal; the Immigration Appeals Tribunal (IAT) remitted a substantial proportion of cases *de novo* to adjudicators; and there was a lack of authoritative and consistent jurisprudence on immigration and asylum cases.⁷⁵ The Government considered that there were two measures necessary to make the system more effective: replacing the current successive rights of appeal with a single right of appeal; and restructuring the appellate authority.

Appeals under the Immigration and Asylum Act 1999

Part IV of the 1999 Act replaces all rights of appeal contained in earlier legislation. The main provisions affecting asylum are those concerning asylum claims (sections 69–73), a new Human Rights Act 1998 appeal (section 65), and the 'one stop' procedure (section 74). Although some consideration was given to replacing the existing two-tier system with a single-tier system, the status quo prevailed. Appeals continue to be heard by the adjudicators (at first instance)⁷⁶ with an appeal to the Immigration Appeals Tribunal.⁷⁷

Section 69 is entitled 'Claims for asylum'. It sets out the rights of appeal on asylum grounds for an individual who: is refused leave to enter (section 69(1)); receives a decision to vary, or a refusal to vary limited leave where the decision would require departure within 28 days (section 69(2)); or receives an order requiring deportation or removal (sections 69(4) and (5)). Such persons can appeal on the basis that the decision would be contrary to the 1951 Convention (section 167). Section 70, in turn, provides for limitations to the section 69 appeal rights. These are mainly in relation to cases in which the Secretary of State feels that exclusion would be in the interest of UK national security. One feature of section 69 is that an appeal cannot take place unless the asylum claim was made *before* leave or variation was refused, or directions made. This limitation is an exact

⁷¹ Asylum and Immigration Act 1996 s 1 amended para 5(5).

⁷² n 11 above, para 9.9.

⁷³ *ibid* para 9.10.

⁷⁴ Home Office/Lord Chancellor's Department, *Review of Appeals – A Consultation Paper* (London, 1998) para 1.1.

⁷⁵ *ibid* para 3.9.

⁷⁶ Note: the Act uses the term 'adjudicator' and there is no longer a differentiation between special adjudicators (who handled asylum appeals) and non-special adjudicators.

⁷⁷ There is now provision in the Act for the President to direct that the IAT may sit with 'such number of members as the President may direct': Sched 2 para 6(3). It is expected that the normal composition will now be two members: Refugee Legal Centre, *Legal Bulletin No: 69* (London: RLC, 2000) 3.

replica of Schedule 2, paragraph 2 to the 1993 Act and was introduced to ensure that an appeal could not lie 'in respect of a decision in relation to an asylum application which [had] not been made'.⁷⁸ In other words, a person cannot bring an appeal on asylum grounds unless an asylum application is lodged before the decision or action is taken against which he is appealing.⁷⁹ For those familiar with UK asylum law, it will be evident that section 69 has many similarities to section 8 of the Asylum and Immigration Appeals Act 1993. As with the 1993 and 1996 Acts, the real bite of the provision is not contained in the main body of the statute but in the Schedules. Thus one has to turn to paragraph 9(2) of Schedule 4 to discover what happens when the Secretary of State certifies a case: where an adjudicator agrees with the certification, there is no further appeal to the IAT. All of the usual circumstances are listed:

- failure to produce a passport without reasonable explanation;
- production of an invalid passport and failure to inform an immigration officer;
- failure to show a fear of persecution on a Convention ground;⁸⁰
- a fear of such persecution is shown, but the fear is manifestly unfounded or the circumstances giving rise to the fear no longer subsist;
- the claim is made after refusal of leave to enter, after recommendation for or notification of deportation, or after notification of removal;
- the claim is manifestly fraudulent, or the evidence in its support is manifestly false;
- the claim is frivolous or vexatious.

As with the 1996 Act, the Schedule contains a torture exemption. Thus sub-paragraph 7 exempts from the certification procedure any claim in which 'evidence adduced in its support establishes a reasonable likelihood that the appellant has been tortured in the country to which he is to be sent'. The one major difference between the 1999 list of certified cases and those contained in the 1996 Act is the notable absence of a White List, which was removed by the Labour government. In contrast to its action on the White List, however, the Government resisted any temptation to comply with its earlier assurances (restoration of a right of in-country appeal),⁸¹ and chose instead to firm up procedures in third-country cases (see above).

Hidden away in a provision entitled 'Miscellaneous limitations on rights of appeal' is a highly significant restriction on appeals (section 72(3)). It states that no appeal may be made in relation to a decision made on an application if the application was required to be made in a prescribed form but was not made in that form, or if the applicant failed to comply with prescribed procedures or time limits. The explanatory notes to the Act fail to offer any reason for the new section; one has to turn for enlightenment to the comments of Mike O'Brien during Special Standing Committee when he introduced the provision as a new government amendment.

The amendment is necessary for two reasons. It supports our application forms policy, which is intended to ensure that in circumstances where the forms are mandatory, applicants assist us in dealing with their cases promptly by providing the necessary information in the required form. It also ensures that a person who does not pursue his application properly should not benefit from a right of appeal if it is refused. No application form is required for asylum or human rights claims.⁸²

⁷⁸ M. Supperstone and D. O'Dempsey, *Immigration. The Law and Practice* (London: Longman, 1994) 374.
⁷⁹ *Ibid* 373.

⁸⁰ Race, religion, nationality, membership of particular social group, or political opinion.

⁸¹ Jane Fiddick, *Immigration and Asylum* (London: Home Affairs Section House of Commons Library, 19 February 1999) Research Paper 99/16, 49.

⁸² HC Special Standing Committee Immigration and Asylum Bill cols 1174–1175 29 April 1999.

Rather surprisingly, in view of its removal of appeal rights for anyone failing to comply with procedural requirements, the amendment was agreed to without debate. Since other appeal rights are protected in the Act, this does seem an unusually restrictive course of action. There is no defence to procedural errors. Consequently, by denying an applicant the opportunity to offer an explanation during an appeal hearing, the Government has once more, and perhaps ill-advisedly, opened up an avenue for judicial review.

The important Immigration and Asylum Appeals (Procedure) Rules 2000 were published in September 2000 and came into force on 2 October 2000.⁸³ They prescribe the procedure to be followed for appeals made under Part IV of the 1999 Act and replace previous appeals rules.⁸⁴ In contrast to the comparable rules issued under the 1996 Act, the new rules provide for a longer period for giving notice of appeal to an adjudicator: not later than ten days after the notice of the decision was received, unless the appeal is made outside the UK, when the time limit is 28 days following departure.⁸⁵ Under the 1996 rules, the appeal time for certain certified cases was two days, an almost impossible time limit to observe.⁸⁶

While the 1999 Act may seem just as restrictive as its predecessor, if not more so, it does have one redeeming feature from a human-rights perspective in that section 65 provides for a new right of appeal based on human-rights grounds. Thus any person lodging a claim to enter or remain in the UK who feels that a decision made by an authority under the Immigration Acts⁸⁷ is in breach of his or her human rights is entitled to appeal to an adjudicator (section 65(1)). The Act explains that an authority acts in breach of a person's human rights when it acts, or fails to act, in a way which is made unlawful by section 6(1) of the Human Rights Act 1998.⁸⁸ For the purposes of the Act, an 'authority' means the Secretary of State, an immigration officer, or a person responsible for the grant or refusal of entry clearance (section 65(7)). Interestingly, the first version of the Bill referred only to an immigration officer or the Secretary of State.⁸⁹ ILPA, notably, was not happy with this restriction on responsible authorities and, in its evidence to the Special Standing Committee, called for an appeal on human rights grounds against decisions made by entry clearance officers in British posts overseas.⁹⁰ The amended version takes account of this request.

The first of the proposals recommended in the consultation document *Review of Appeals*, that of a single right of appeal, is provided for in sections 74–78. They require the decision-maker to serve a notice on any person appealing a decision, calling for any additional grounds he or she may have to enter or stay in the UK. The aim is that all additional grounds will be considered alongside any refusal of the asylum claim.⁹¹ Any failure to reveal a particular ground will mean that the appellant will be unable to rely on it at appeal, except where the ground is an

⁸³ SI 2000 No 2333.

⁸⁴ Immigration Appeals (Procedure) Rules 1984; Immigration Appeals (Procedure) (Amendment) Rules 1991; Immigration Appeals (Procedure) (Amendment) Rules 1993; and Asylum Appeals (Procedure) Rules 1996.

⁸⁵ Rules 6(1) and (2)(a).

⁸⁶ Note, however, the 2000 Rules cater for transitional arrangements (Rule 4(4)) and the accelerated time-limits therefore continue to apply to appeals made under s 8 of the 1993 Act and s 3 of the 1996 Act.

⁸⁷ Excluding the British Nationality Act 1981.

⁸⁸ s 65(2). Human Rights Act 1998 s 6(1) states: 'It is unlawful for a public authority to act in a way which is incompatible with a Convention right.'

⁸⁹ Immigration and Asylum Bill (Bill 42) clause 47.

⁹⁰ HC Special Standing Committee Immigration and Asylum Bill col 218 17 March 1999 (ILPA).

⁹¹ Explanatory Notes to Immigration and Asylum Act 1999 para 219.

asylum claim or a claim that an act breached the applicant's human rights, or that the Secretary of State believes that the applicant had a reasonable excuse for the omission (section 76(3)). However, any applicant who tries to claim asylum after the period prescribed will have no right to an appeal against refusal if the Secretary of State certifies that one of the reasons for making the asylum claim was to delay removal from the UK and that there was no other legitimate purpose for making the application (section 76(5)). Regulations explaining further the procedure to be adopted in one-stop appeals have now been published.⁹²

The second proposal, that of a restructured appellate authority, is dealt with in Schedules 2 and 3 to the Act. Though the original Bill was a little more adventurous in its proposals, such as removing 'lay' members from appointment to the IAT,⁹³ the final version makes few major changes. The most noteworthy is the power granted to the President of the IAT to appoint such number of members as he or she thinks fit (the previous requirement was three) (Schedule 2 para 6(3)). It is clear that jurisdiction may be exercised by a single member, although it is now felt that this is unlikely to be the case for substantive appeals.⁹⁴ With regard to the adjudicators, there is little change, although certain matters are written into statute for the first time, such as the appointment of the Deputy Chief Adjudicator (Schedule 3 paragraph 1). Furthermore, there is no longer any provision for the appointment of 'special' adjudicators and, therefore, all adjudicators, whether handling asylum cases or not, will be known simply as 'adjudicators'.

One final point of interest added by the 1999 Act is the power granted to the IAT to notify an appellant, at any time prior to determining an appeal, that it considers that the appeal has no merit (section 79(1)). The IAT may then go on to impose a financial penalty, as specified by order by the Lord Chancellor (sections 79(5) and (6)). This is again a measure aimed at deterring 'unnecessary' appeals and relieving the pressure on the appellate bodies.

Detention and bail

Possibly more than any other area of asylum/immigration policy, detention has succeeded in becoming an issue of public concern. Long-running campaigns, such as that at Campsfield Detention Centre near Oxford, have heightened public awareness of detention under the Immigration Act 1971 and the wide powers vested in immigration officers to detain.⁹⁵ Despite calls for an end to detention, and its description as a 'national scandal',⁹⁶ the Government reaffirmed its commitment to immigration detention in the White Paper: 'Effective enforcement of immigration control requires some immigration offenders to be detained' and 'detention is necessary to ensure the integrity of our immigration control'.⁹⁷ This despite an internal Home Office review of detention carried out in August 1997, and the critical reports of Sir David Ramsbotham, HM Chief Inspector of Prisons, on the Tinsley House and Campsfield detention centres published in April 1998.

In accordance with this faith in the principle of detention, the Government introduced only minor changes to the provisions on detention. Part VIII of the 1999 Act places the current system on a statutory footing, making provision for the

⁹² Immigration and Asylum Appeals (One-Stop Procedure) Regulations 2000 SI 2000 No 2244.

⁹³ Bill 42 Sched 2 para 1(3).

⁹⁴ Refugee Legal Centre, n 77 above.

⁹⁵ Immigration Act 1971 Sched 2 para 16.

⁹⁶ *Briefing on the Immigration and Asylum Act 1999* (London: Refugee Council, 2000).

⁹⁷ n 11 above, paras 12.1 and 12.3.

management and operation of detention centres, and setting out the responsibilities of detainee custody officers. An innovation is contained in section 153, which requires the Secretary of State to make rules for the regulation and management of detention centres. These rules have now been published, and are based on prison rules (Detention Centre Rules 2001, SI 2001 No 238).

While the attempt to introduce greater control over detention facilities was welcome, the most important development was in relation to bail. The pre-1999 bail position was much criticised. As indicated above, under the Immigration Act 1971, immigration officers have a wide power of detention. This was not matched by an equivalent power of release. Schedule 2 to the 1971 Act sets out certain rules with regard to release and bail, which were further developed by changes made under the Asylum and Immigration Appeals Act 1993 and the Asylum and Immigration Act 1996. Simply stated, a detainee is entitled to apply for bail from a chief immigration officer (or higher) or from an adjudicator so long as seven days have elapsed since the detainee's arrival in the UK, and a recognisance is entered into.⁹⁸ In the past, the recognisance necessitated the production of one or more individuals of good character and standing, who had evidence of their means, and were prepared to act as sureties to a named sum. The detainee also had to provide a place of residence on release, preferably with one of the sureties. The sum demanded was left to the discretion of the adjudicator or officer in charge of the proceedings, but it was normal practice in the early 1990s to ask for £1,000 minimum from each of two sureties.⁹⁹ All detainees were treated alike, whether asylum seekers or not.¹⁰⁰

The 1999 Act introduced an important new element to the bail procedures provided for by the Immigration Act 1971: the provision for two routine bail hearings, the first taking place no later than the tenth day following the initial detention, the second no later than the thirty-eighth day following detention (sections 44(7)(b)(i) and (ii)). These new routine bail hearings now apply to all those detained under the Immigration Act 1971, subject to a few exceptions: those detained under a provision other than the 1971 Act; those liable to deportation as a result of a court's recommendation (s 3(6) of the 1971 Act); and those who have informed the Home Secretary in writing that they do not wish to be considered for routine bail hearings (section 44(3)). These changes now apply alongside existing procedures and do not replace them.

In addition to the new right granted to detainees, section 46 introduced a presumption in favour of bail, bringing immigration bail provisions in line with the Bail Act 1976 for the first time. Clearly, the presumption in favour of bail is a highly significant innovation and one which was unexpected since the original Bill contained no such presumption. During Special Standing Committee, the NGOs argued strongly in favour of a presumption of liberty akin to that existing under criminal law. One reason proposed was compliance, or non-compliance, with the European Convention on Human Rights. Justice, for example, provided evidence that a failure to incorporate a presumption in favour of bail would bring the Act into conflict with Article 5 of the ECHR (right to liberty) as well as Article 9 of the International Covenant on Civil and Political Rights (ICCPR).¹⁰¹ The Government

⁹⁸ Immigration Act 1971 Sched 2 para 22.

⁹⁹ For a copy of the practice note see *Butterworths Immigration Law Service*.

¹⁰⁰ See for further detailed discussion of detention practice in the 1990s Amnesty International, *Prisoners without a Voice: Asylum Seekers detained in the United Kingdom* (London: Amnesty International, 1995).

¹⁰¹ See HC Special Standing Committee Immigration and Asylum Bill col 65 16 March 1999 (Justice).

partly conceded the point and included an amendment allowing for a 'General right to be released on bail'.

A further point of interest regarding bail is the provision on location of bail hearings (section 45). This enables the Secretary of State to direct where a bail hearing may take place. Thus hearings may be held not only in a court but also, controversially, in prisons or in detention centres, thereby denying the detainees a hearing in public. ILPA, in particular, objected to the lack of public access in such cases, believing that matters 'relating to the freedom of an individual must be open to the public (unless privacy has been requested), and must be seen to constitute an independent review'.¹⁰² Rather bizarrely, the Act also permits the use of live television links at bail hearings, whereby the detainee remains at the detention centre/prison and his or her representative is present in court. Not only will the legal representative have difficulty in taking instructions during such hearings, but the use of live link will undoubtedly place the detainee at a disadvantage in giving evidence.¹⁰³ ILPA also considers that it is of paramount importance that a detainee is produced when liberty of the individual is under review and that failure to do so 'will lead to the objectification of the detainee'.¹⁰⁴ While there is much to commend the new bail provisions, there is also much that continues to raise concerns for rights-focused lawyers and NGOs.

Support

The pre-1999 support system

While the polarisation of views on asylum is unsurprising – particularly in relation to rising asylum applications, appeals, carriers' and employers' liability – differences of opinion are most pronounced where financial and practical support for asylum seekers is concerned. Successive governments have been aware of their obligations under international human rights law and reluctant to adopt unnecessarily draconian measures. Nevertheless, some NGOs would argue that recent initiatives have come close to creating a new destitute class among waiting refugee claimants. For example, the Asylum and Immigration Act 1996 limited certain social security benefits to those who applied immediately on arrival at ports, or who were subject to a 'state of upheaval' declaration.¹⁰⁵ Housing for asylum seekers was also restricted with the result that many found themselves reliant on non-state sources of accommodation. These changes in the law led directly to appeal in the courts. In a widely discussed decision handed down on 8 October 1996, *R v Hammersmith and Fulham LBC, ex parte M*, Mr Justice Collins delivered a stinging indictment of government policy. He held that a destitute asylum seeker who had neither money nor means of support could be in 'need of care and attention' within the meaning of section 21(1)(a) of the National Assistance Act 1948. Collins J refused to accept that Parliament had intended that an asylum seeker, who was lawfully in the UK, should be left destitute, starving and at risk of grave illness and even death because he could find no one to provide him with the bare essentials of life. Fearing a rising financial burden as a result of this judgment, the appellant councils challenged the decision in the Court of

¹⁰² HC Special Standing Committee Immigration and Asylum Bill col 212 17 March 1999 (ILPA).

¹⁰³ *ibid* col 213.

¹⁰⁴ *ibid*.

¹⁰⁵ Asylum and Immigration Act 1996 s 11; a state of upheaval declaration applies to individuals who are in the UK but whose countries the Secretary of State decides are in a state of upheaval, thereby giving them an opportunity to apply for asylum.

Appeal, only to be turned down. Consequently, and in recognition of the burden upon local authorities, the Government was forced to pay them special accommodation grants, amounting to several million pounds in total.¹⁰⁶

While the incoming Labour government admitted that cash-based support was administratively more convenient and usually less expensive to administer,¹⁰⁷ it also recognised that the support arrangements had become untenable.¹⁰⁸ After due consideration, the Government proposed some surprisingly radical solutions: the removal of all rights to welfare benefits for asylum seekers; the amendment of the National Assistance Act 1948 to ensure that responsibility for destitute asylum seekers no longer fell upon local authorities; the removal from local authorities of any outstanding obligations under homelessness legislation for asylum seekers; and the introduction of new 'support in kind' arrangements run by a national agency.¹⁰⁹ Local authorities would, however, remain responsible for unaccompanied children and families with children under the Children Act 1989.

Support arrangements under the Immigration and Asylum Act 1999

Despite widespread criticism, and on-going discussions between the NGOs and the Home Office, the new proposals were introduced in the Bill and sanctioned under the 1999 Act. A key factor behind government thinking was the cost of the pre-1999 scheme¹¹⁰ and it remained unconvinced that an alternative was more cost-effective. The structure of the new scheme provided by the 1999 Act is as follows: all entitlement to benefits is ended for all 'persons subject to immigration control' which includes asylum seekers (section 115). In its place is a new support mechanism run by the National Asylum Support Service (NASS) based in the Home Office. The Home Secretary is granted a power to support asylum seekers; he has no duty to do so (section 95(1)). The Act spells out which persons are entitled to support, namely asylum seekers and their dependants where they appear to the Secretary of State to be either destitute or likely to become destitute within a prescribed period (normally 14 days).¹¹¹ The parties concerned must make an application for support and must meet the criteria for destitution set out in the Act and in secondary legislation: the Asylum Support Regulations 2000. Destitution is defined as not having adequate accommodation or any means of obtaining it (whether or not essential needs are met), or having adequate accommodation, but not having essential living needs (section 95(3)). The scope of support includes accommodation, if required, vouchers and some cash (£10 per person or £20 per qualifying couple per week).¹¹² The Regulations set out the amounts which differing categories of asylum seeker may expect to receive by way of vouchers (including those redeemable for cash). Thus, for example, a qualifying couple is entitled to £57.37 per week; a single person aged 25 or over, £36.54; a single

106 See for example HL Deb vol 581 WA col 158 22 July 1997 for cost incurred by the Government from 1996-7: £5.6 million. In the White Paper, the Home Office declared support arrangements to cost about £400 million per year: n 11 above, para 8.14. From 1998 to 1999, the funding for a single adult asylum seeker was increased by almost 18 per cent to £165 per week to cover accommodation: HL Deb vol 586 col 403 23 February 1998.

107 Fiddick, n 81 above, 56.

108 Mike O'Brien speaking at Immigration Advisory Service Annual Conference, Sheffield, 18 July 1997.

109 See n 11 above, paras 8.12-8.26.

110 See Home Office Consultation Paper, *Asylum Seekers Support - An information document setting out proposals for the new support scheme for asylum seekers in genuine need and inviting expressions of interest from potential support providers* (1999) preface.

111 s 95(1); The Asylum Support Regulations 2000 SI 2000 No 704 reg 7.

112 *ibid* reg 10(6).

person aged 18–25, £28.95; and a child under 16, £26.60.¹¹³ These amounts are equivalent to 70 per cent of income support rates.

With regard to the adequacy of accommodation provided, section 95(6) excludes certain matters from consideration when an offer of accommodation is made.¹¹⁴ Section 96 sets out the ways in which the support may be provided (accommodation, essential living needs, expenses (other than legal) associated with the asylum claim), and section 97 requires the Secretary of State to have regard to certain factors when providing accommodation under section 95.¹¹⁵ Since most asylum seekers would probably prefer to remain in London near their own communities, the Secretary of State is specifically excluded from taking into account any preference of the asylum seeker as to locality of the accommodation (section 97(2)). Instead, asylum seekers will be dispersed to towns and cities around the country. The Act also leaves open further exclusions by order; these have now been identified as any personal preference as to the nature of the accommodation to be provided or the fixtures and fittings.¹¹⁶ Asylum seekers are not permitted to be absent from their authorised address for more than seven consecutive days and nights without permission, nor must they be absent for more than 14 days and nights in any six month period.¹¹⁷ If an asylum seeker's application for asylum is met with a positive decision (refugee status or ELR), he or she will have only 14 days' grace in which to look for other means of support and housing, since both will cease after that period.¹¹⁸ Refugees or those with ELR will, however, be eligible for welfare benefits and, if in 'priority need', will be eligible for local authority housing. The above restrictions might tempt some to refuse the offer of housing. However, they should be warned that any such refusal will result in their having to make their own arrangements for accommodation. Some asylum seekers may, of course, have alternative housing provision, and in these circumstances they may apply solely for support.

True to its word, the Government amended section 21 of the National Assistance Act 1948 to exclude its application to persons subject to immigration control; this no longer exists as a loophole for potentially destitute asylum seekers (section 116). The one category of asylum seeker which continues to be handled differently is that of unaccompanied children under 18 years of age; they alone remain the responsibility of local authority social service departments under the Children Act 1989.

The 1999 Act has also established an appeal mechanism for asylum seekers who have been refused support or accommodation by NASS. Section 102 of and Schedule 10 to the Act create the new office of Asylum Support Adjudicator (ASA). The structure will be similar to that of normal adjudicators in that there will be a Chief Asylum Support Adjudicator, a Deputy, and a number of other ASAs. They will hear appeals against refusal or cessation of support. If an ASA finds in favour of an appellant, he or she may require the Secretary of State to reconsider the decision, or may impose his or her own decision. If the applicant fails in the

113 *ibid* reg 10(2).

114 These are: (a) the asylum seeker has no enforceable right to occupy the accommodation; (b) the accommodation is shared with others; (c) the accommodation is of a temporary character; and (d) the location of the accommodation.

115 These are: the fact that the accommodation is to be temporary; the desirability of providing accommodation in areas in which there is a ready supply of accommodation; and any other matters which may be prescribed.

116 s 97(2)(b); The Asylum Support Regulations 2000 SI 2000 No 704.

117 *ibid* reg 20(e).

118 *ibid* reg 2(2).

appeal, there is no further recourse to the courts, although judicial review remains a possibility.¹¹⁹

The new arrangements were severely criticised by refugee-support groups. Early in the history of the Bill, NGOs campaigned energetically against vouchers and in favour of cash benefits. The Government held firmly to its view: cash was a pull-factor for asylum seekers and had led to the sudden surge in applicants in recent years.¹²⁰ This opinion was countered repeatedly by organisations such as the Refugee Council, which argued that 'there is no real evidence to support the contention that benefits attract and vouchers deter.'¹²¹ A number of Labour backbenchers also disagreed with the voucher system and, consequently, late in the debates on the Bill, Neil Gerrard MP, a member of the Special Standing Committee, tabled an amendment which flew in the face of the Government's support proposals: he called for the restoration of benefits and the withdrawal of vouchers. Smarting from wounds inflicted by 67 rebel MPs over the Welfare Bill, and fearing a further revolt, the Government made a number of hasty concessions: the promise of a part cash payment to each asylum seeker of £10 maximum per week (from £7 a week for adults and £3.50 a week for children), the promise that families with children would not have to wait more than six months on vouchers after April 2000, and a pledge to inject a further £60 million into the decision-making process.¹²² This seemed to suffice and the amendment was withdrawn, the sole remaining option for dissenting Labour MPs being to vote against the entire Bill. In the event, only seven (heralded as the 'Magnificent Seven' by the Joint Council for the Welfare of Immigrants (JCWI)) entered the 'No' lobby with the Liberal democrats,¹²³ while the Conservatives abstained.

This partial victory has not mitigated the unease with which vouchers are regarded. Not only are they considered demeaning, but shops are prevented from giving change to asylum seekers using vouchers. From the date on which the voucher scheme came into operation, 3 April 2000, a number of charities – Oxfam, Save the Children Fund, Shelter and Marie Curie Cancer Care – decided not to participate in the scheme on the basis that refusing retailers permission to provide change for vouchers was unreasonable and discriminated against asylum seekers when they were at their most vulnerable.¹²⁴ In addition the Asylum Support Regulations 2000, expressly excluded toys from the definition of essential living needs. It was argued that the regulations were incompatible with both the UN Convention on the Rights of the Child and the ECHR.¹²⁵ The Home Office countered this claim, stating that 'the purpose of the provision is to ensure that families are not regarded as destitute simply because they cannot afford toys for their children.'¹²⁶ There was, apparently, no intention to prevent voluntary and community organisations from providing toys for the children of asylum seekers. However, regulations 6 and 12 allow the Secretary of State to take into account any charitable contributions made in cash to asylum seekers. Such measures may arguably deter members of the public from making donations to certain charities. As Baroness Williams put it: 'Which of us is going to give to a refugee fund if we

119 s 103. See also *The Asylum Support Appeals (Procedure) Rules* 2000 SI 2000 No 541.

120 For example, see HC Deb vol 326 col 46 22 February 1999.

121 *The Refugee Council*, n 96 above, 11.

122 'Alan Travis, 'Straw moves to quell asylum bill rebellion' *The Guardian* 9 June 1999.

123 *JCWI Bulletin* (London: Summer 1999) 7.

124 HL Deb vol 612 col 895 20 April 2000.

125 See for example Baroness Williams, HL Deb no 1804 col 908 20 April 2000.

126 Cited in HL Deb vol 612 col 896 20 April 2000.

know that that will simply come off levels of support?'¹²⁷ The feeling aroused by the new support scheme was so strong that Lord Dholakia called for annulment of the support regulations in the House of Lords and a debate was consequently held.¹²⁸ While it was not expected actually to lead to annulment, and was viewed simply as 'a way of stamping one's feet',¹²⁹ it was an opportunity to air grievances and stands as a useful record of the issues at stake in the asylum support debate.

On the matter of accommodation, the NGOs spoke out against the one-off offer and dispersal of asylum seekers around the country. Citing previous examples of dispersal – the Vietnamese in the 1970s and 80s, and the Bosnians in the 1990s – they pointed out that dispersal had a chequered history. Either such refugees had faced a risk of harassment from indignant locals, or they had drifted back to towns where there was a well-established community from their own country.¹³⁰ In the short time during which interim dispersal measures were implemented voluntarily by the Local Government Association (December 1999–April 2000), it was clear that dispersal was not a simple process to organise. Offers of accommodation by the regional consortia had not been as forthcoming as anticipated, leading to a relatively low-level of dispersal: only 1,910 cases against an original target of 10,400 by 31 March 2000.¹³¹

The new support scheme was due to come into effect from 3 April 2000 across the whole of the UK but was in fact delayed in England and Wales. Port applicants claiming after 2 April 2000 entered the scheme, while in-country applicants remained the responsibility of the social services departments for a short period and were subject to a NASS roll-out programme. From 25 September 2000, no local authority in England, Wales, Scotland or Northern Ireland should have responsibility for new asylum seekers, or for those from whom social security benefit entitlement has been withdrawn.¹³² In June 2000, the Audit Commission published a detailed report on the implementation of dispersal under the Immigration and Asylum Act 1999.¹³³ It concluded with a number of recommendations to local authorities and regional consortia, and to government departments. Information-gathering, good management and clear communication between the relevant bodies were highlighted as needing improvement. Constant review of asylum policy and provision for asylum seekers was also advocated.

While the Refugee Council has admitted that the NASS system is better than that run previously by local authorities, it continues to gather information which confirms the Audit Commission's conclusions. Some asylum seekers are beginning to drift back to London, preferring to sleep on floors than to remain in their NASS housing outside the capital. There are reports of poor accommodation and unnecessary charges being levied against asylum seekers. Certain providers have failed to assist with access to legal advice, as well as to the essential services of education and health,¹³⁴ and, in the worst cases, harassment of asylum seekers has occurred. As predicted, the voucher scheme is proving problematic with asylum seekers only able to redeem their cash vouchers in post offices and unable to obtain

127 *ibid* col 909.

128 HL Deb vol 612 cols 893–920 20 April 2000.

129 *ibid* col 902.

130 Audit Commission, *Another Country – Implementing Dispersal under the Immigration and Asylum Act 1999* (London: Audit Commission, 2000) para 20.

131 *ibid* para 21.

132 Home Office IND, 'National asylum support service: roll-out of support scheme' 10 July 2000.

133 Audit Commission, n 130 above.

134 Refugee Council, *In Exile* August 2000 2.

change from transactions, with the result that shops accepting vouchers continue to profit at the expense of asylum seekers. Furthermore, there is evidence to suggest that a black-market is developing in vouchers, since asylum seekers exchange them for cash wherever possible, but may only obtain about 70 per cent of their value.¹³⁵ Having had its pensions policy voted against at the Labour Party Conference in September 2000, and fearing further problems on asylum policy, the Government, through its Immigration Minister, Barbara Roche, announced to the Conference on its last day that asylum seekers would henceforth be given change on using vouchers.

Conclusion

The tiresome mantra 'fairer, faster, firmer', uttered on every occasion by Ministers when discussing the White Paper or Bill, provides a useful benchmark against which the Immigration and Asylum Act 1999 may be measured. For an indication of the Government's own definition of the terms 'fair, fast and firm', one can turn to Jack Straw's own words during the Second Reading of the Bill:

We want a fairer system that reflects our commitment to race equality and human rights; we want a faster system that is able to deal quickly with all applicants, whether visiting this country or seeking to remain here longer; and we want a firmer system, with strong control at ports and effective enforcement against those not entitled to stay.¹³⁶

In the foregoing discussion of the Act, serious doubts have been raised on a number of counts, in particular that of human rights compliance. Many of the provisions have been held to be falling short of the UK's international obligations, particularly the ECHR. The one saving grace appears to be the introduction of a human rights appeal, yet even here problems may emerge. Recourse to this right is likely to be exercised more frequently than the Government may have anticipated and the inevitable consequence will be further delays in the courts. Any system which encounters lengthy delays in its appellate process will face accusations of unfairness and possibly of weakness. On the other hand, 'fast' controls too may be deemed unfair. Even the former Chief Adjudicator and President of the IAT, Judge David Pearl, speaking in 1997, questioned whether fairness and speed were compatible objectives, and warned that adjudicators had to ensure that their decision-making, if fast, was also consistent to ensure fairness.¹³⁷ One change incorporated under the Act should improve the appeals process: the one-stop appeal. Yet this new arrangement does not remove the right to seek judicial review which remains a viable option in many circumstances. Both preceding Acts of 1993 and 1996 fell into a similar trap; as one door closed on asylum appeals, another opened for judicial review. Again, there is likely to be an unwelcome impact on the court lists.

Fairness in the appeals procedures must also be accompanied by fairness in initial decision-making. Many called for an injection of capital to help overcome the Home Office's chronic understaffing and to address the alarming pile of undecided asylum cases. Following a catastrophic year in 1999, when applications peaked and the backlog grew exponentially, the Home Office has shown some signs of improvement, despite admitting that it had lost over 8,000 immigration

¹³⁵ Bill Morris speaking at the Labour Party Conference 28 September 2000.

¹³⁶ HC Deb vol 326 col 37 22 February 1999.

¹³⁷ Speaking at the IAS Annual Conference, Sheffield, 18 July 1997.

casework files.¹³⁸ Initial decisions have started to rise from the derisory 2,320 recorded in December 1999 to a high of 11,340 in March 2000 (though July 2000 saw a drop once more to 8,855).¹³⁹ If the Government ever reaches its target of two months to decide each new case and a further four months for determination of the appeal, then progress may be made. This is, of course, dependent on good, consistent decisions at all stages of the process.

Where the Act does meet its expectations is in relation to firmness; yet it remains unclear whether the huge enforcement arsenal contained in the legislation will be able to contend with the growing crime of human smuggling. The probable answer is that it will not. Transporters of all types are facing an increasingly difficult future as they attempt to meet the stringent requirements of the new laws. As the Deputy Managing Director of Eurotunnel conceded, 'No transport system in the world is 100% able to ensure illegal travellers don't board.'¹⁴⁰ Migrants continue to exercise every available means by which to gain entry to Europe, some with tragic consequences; one such saw at least 13 men jump off a freight train outside Folkestone, an episode which resulted in serious injuries.¹⁴¹

A fair asylum policy must be one which is clear and comprehensible. The 1999 Act suffers from unusual length and an inordinate reliance on secondary legislation. While several statutory instruments have now emerged, there are many still outstanding. A full understanding of the Act is thus delayed and an impression of overly technical legislation fostered. The Government might choose to counter accusations of unnecessary complexity with the claim that it has introduced a much-needed revolution of asylum policy within the wider context of migration to the UK. Unable to extract itself from the inevitable 'economic migrant versus refugee' debate, the Government was forced to repeat or expand many of the formulae of previous legislation. A truly innovative approach would have addressed the causes of increasing migratory flows and re-opened a discussion of the advantages of legal migration, as past exemplified by Canada and Australia. Such a course is generally viewed as political suicide. Nonetheless, in a highly significant move in July 2000, the Home Office announced a new immigration scheme to attract entrepreneurs with innovative business ideas to the UK.¹⁴² This was an admission that not all economic migrants are a burden on the country and that, in fact, they may promote substantial economic benefit. The importance of such a development should not be overlooked. More will be required, however, if the UK is to redress the balance of an asylum policy which has tipped increasingly towards exclusion and which culminated in the Immigration and Asylum Act 1999.

138 Refugee Council, n 134 above.

139 Home Office, *Asylum Statistics July 2000*.

140 Keith Perry, 'Eurotunnel immigrants injured' *The Guardian* 2 September 2000.

141 *ibid.*

142 Home Office Press Notice 219/2000 25 July 2000.

REPORTS

Reforming Directors' Duties

*Sarah Worthington**

Taken together, the English and Scottish Law Commissions,¹ the Department of Trade and Industry (DTI)² and the Company Law Review Steering Group (CLRSG)³ have produced hundreds of pages on the subject of directors' duties during the last three years. All of this work is just a small part of the DTI's current review of the whole system of core company law, a review billed as the most comprehensive ever undertaken in the United Kingdom (UK). This note cannot do justice to the detailed work of all these bodies, but aims simply to summarise the principal directions of the work so far produced – and only so far as it relates to directors' duties – and to highlight some aspects which give cause for concern. In doing that, it is consciously biased towards consideration of the perceived problems, rather than praise for noteworthy advances. The reason is simple. This mammoth review process is not yet completed. The CLRSG is due to produce its final report in Spring 2001. Before that date, it seems important to cast a critical eye over the work completed so far. That said, it would be surprising if the CLRSG did not already intend to address some of the issues raised here. Its work to date has always been presented as part of a consultation process aimed at discovering the optimum solution.

The need for a comprehensive reform of UK company law is not doubted. The UK rules have long ceased to provide an enviable model for other countries to adopt. The current companies legislation is widely regarded as being too complex and detailed, and as containing rules which are now either obsolete or unwarranted.⁴ The DTI's stated aim is to produce a simple, rational framework which is modern and competitive, and which facilitates enterprise and promotes transparency and fair dealing.⁵ It is against this that its efforts need to be judged.

* London School of Economics and Political Science; member of the Company Law Review Working Group on Sanctions. The views expressed here are personal, and do not reflect the views of this Working Group.

1 Law Commission, *Company Directors: Regulating Conflicts of Interest and Formulating a Statement of Duties* (1998, Law Com No 153, Scot Law Com No 105) ('LCCP'); Law Commission, *Company Directors: Regulating Conflicts of Interest and Formulating a Statement of Duties* (1999, Law Com No 261, Scot Law Com No 173) ('LCR').

2 DTI, *Modern Company Law for a Competitive Economy* (1998) ('MCL').

3 CLRSG, *Modern Company Law for a Competitive Economy: The Strategic Framework* (1999, URN 99/654) ('SF'); CLRSG, *Modern Company Law for a Competitive Economy: Developing the Framework* (2000, URN 00/656) ('DF'); CLRSG, *Modern Company Law for a Competitive Economy: Completing the Structure* (2000, URN 00/1335) ('CS').

4 MCL 6–7.

5 MCL Foreword and para 3.8.

Summary of critical conclusions

The CLRSG recognises that the issue of directors' duties lies at the heart of corporate governance.⁶ It says its proposals 'go a long way towards sketching out a radically improved framework of company law for the new Century.'⁷ True, its proposals involve major changes to the legal presentation of directors' duties (a statutory enactment of both the existing common law duties and their remedies is recommended), but they recommend little conscious change to the substance. Perhaps the most controversial and hard fought debate relates to the 'scope' issue: for what purposes and in whose interests should companies be run? The CLRSG concluded that the overall objective of wealth generation and competitiveness for the benefit of all could best be achieved through adopting an 'inclusive' or 'enlightened shareholder value' approach, rather than a 'pluralist' or 'stakeholder' approach.⁸ The former approach puts shareholders at centre stage: companies are to be run primarily for the benefit of shareholders, although directors are required to take relevant decisions in an 'inclusive' fashion, so as to recognise best practices in relation to a broad range of interested groups (be they employees or customers, or the wider community which might be affected by social or environmental concerns). The 'pluralist' view, on the other hand, would give shareholders no inevitable primacy, but would require the directors to balance the interests of all the relevant stakeholder groups in determining whether a course of action was in the 'interests of the company'. This issue is taken up later.

In the end, this inclusive, enlightened shareholder value regime appears to be distinctly pluralist in objective, even if avowedly traditionalist in substance. This view is reinforced by the arguments advanced for adoption of a related strategy, the imposition of an enhanced reporting regime for listed and large private⁹ companies.¹⁰ This new report (part of the Annual Report), called the Operating and Financial Report (OFR), is designed to be a full account by the directors of the performance and direction of the company's business, including a review of achievements, trends and strategic direction. In addition, however, the OFR is to include a report on the company's wider relationships, risks and opportunities, and social and environmental impacts, where these are relevant to an understanding of the business.¹¹ This proposal for a broad-ranging mandatory OFR seems to mark a formal recognition of the increasing interest of a wide range of stakeholders in the greater public disclosure of corporate functions and activities.¹²

Otherwise little is to change. There is no recommendation for institutional changes in board structure (often seen as a necessary component of a pluralist governance regime).¹³ There is no discussion of the divided loyalty problems experienced by both nominee directors and directors of companies forming part of a group. There is scant discussion of the relationship between directors and third

6 DF para 3.2.

7 DF para 1.8.

8 Much has been written on this debate. See, for example, J. Parkinson, *Corporate Power and Responsibility* (Oxford: OUP, 1993); P. Ireland, 'Corporate Governance, Stakeholding and the Company: Towards a Less Degenerate Capitalism' (1996) 23 *Journal of Law and Society* 287; J. Kay and A. Silberston, 'Corporate Governance' (1995) *National Institute Economic Review* 84; F. Easterbrook and D. Fischel, *Economic Structure of Corporate Law* (Boston: Harvard, 1991) ch 1.

9 An annual turnover in excess of £500 million is suggested: CS para 3.2.

10 DF paras 5.79 ff; CS ch 3.

11 SF paras 2.16–2.18; CS paras 3.1–3.4, 3.32–3.42.

12 eg DF para 2.21.

13 DF paras 3.138–3.141, although also see para 2.24 on enabling structures.

parties,¹⁴ directors' training, qualifications, and terms and conditions,¹⁵ the role of non-executive directors,¹⁶ and the overall functioning of the corporate governance system.¹⁷ All the consultation papers are alert to self-regulatory regimes and codes of best practice which are in place for listed companies, but these options have not impacted on proposals for core company law reform.¹⁸ Finally, so far as directors' duties are concerned, there has been little obvious concern with the specific issues faced by directors of small companies, despite the DTI's assertion that it will 'think small first'. Initial suggestions for a different governance structure for owner-managed companies¹⁹ have been shelved as inimical to transitions to larger company structures.²⁰ In short, the proposals generally reflect existing rules; they do not purport to design afresh for modern conditions. Perhaps this tried and tested regime does provide the best possible framework, but little evidence is offered to support that conclusion.

Chronology of reform – the importance of the scope issue

At a very fundamental level it might be suggested that this conservative result was inevitable because the plan of attack in the law reform process was compromised from the start. A sensible discussion of corporate governance cannot take place without adopting a position on the role of companies: what interests should companies, and company law, serve? Only then is it possible to assess which duties ought to be imposed on directors and, after that, how and by whom those duties can best be enforced. By accident rather than by design, the current process has addressed these issues in precisely the reverse order, commencing with an examination of certain limited enforcement issues²¹ and leaving assessment of the function of modern corporations (the 'scope' issue) to the concluding stages. Even in the final stages, the reform perspective has been subtly distorted: the CLRSG has not considered itself able to deal with insolvency-related issues for fear of transgressing into territory reserved to the contemporaneous review of insolvency matters. Maybe no great harm has been done, but it is hard to imagine that certain issues and possibilities have not been missed.

On the specific issue of directors' duties, for example, the CLRSG adopts much of the work of the Law Commissions. This is to be expected, yet the Law Commissions' task was to deal simply with the presentation of the general law, not its reform.²² Their remit did not permit them to consider what interests companies ought to serve, the relevance of corporate size, or the possibilities of alternative board structures; moreover, their final report did not deal with the role and function of different types of directors (executive, non-executive, nominee, shadow or alternate).²³ Later, when the CLRSG came to determine the critical 'scope' issue (ie in whose interests should companies be operated and controlled?), it simply backed its conclusions into the Law Commissions' findings. It used its answer to

14 DF paras 3.92–3.104; CS para 4.22ff.

15 DF paras 3.105–3.111; CS paras 4.42–4.43.

16 DF paras 3.132–3.153; CS paras 4.44–4.47.

17 DF paras 3.126–3.128 and 3.154–3.168.

18 LCR para 1.11; DF paras 3.112–3.121.

19 DF paras 7.95–7.107.

20 CS paras 2.35–2.36.

21 Law Commission, *Shareholder Remedies* (1997, Law Com No 246).

22 LCR para 1.3.

23 LCR para 1.32.

the scope question simply to interpret and reframe the traditional duty imposed on directors to act *bona fide* and in the interests of the company.

Had the work started with the scope question, then maybe the proposals would have been more ambitious. Starting with issues of scope, it might have been possible to step back from company law as traditionally formulated, and to consider, in the context of a discussion of the purpose and functions of companies, what protections the law ought to provide, and to whom, and how those protections might best be formulated and enforced. For example, there are already many constraints on what a company can do. These include general constraints (eg employment law), company-specific constraints (eg capital maintenance rules), and constraints directed at specific companies (eg specific objects clauses). Who are all these rules designed to protect? Is it the company's shareholders, or some broader community of interests? Should primary liability for failures rest with the company or the company's directors? If the latter, is enforcement best mediated by penalties and criminal law sanctions, or by civil law actions for damages? Can these answers be rationally tailored to meet the reform objective that directors' duties be stated comprehensively and in an accessible form? Much the same questions can be asked in relation to the disclosure obligations imposed on companies and their directors, and indeed also in relation to the specific common law and equitable duties imposed on directors. Instead, many of the current proposals seem manacled to their historical roots; the focus appears to be bottom up, rather than top down, with the result that proposals may be unduly conservative.

Guiding principles of reform

Despite these problems with chronology, most observers probably still expected the reform proposals to be quite radical. The entire reform process has been governed by carefully articulated principles designed to achieve imaginative ends. This consciously self-reflective process began with the early Law Commissions' work on shareholder remedies.²⁴ There the Law Commissions set out a long list of 'guiding principles', twelve in all, against which reform measures could be evaluated. These principles included the idea that the law should be a facilitator, working efficiently and cost-effectively, without imposing excessive regulation, and generally without interfering with contractual arrangements or proper commercial judgements. The law should be accessible, comprehensible, clear and consistent with common sense, and, wherever possible, it should also be precise and certain. Sanctions should be appropriate. Within companies, shareholders and directors should have separate but interdependent roles, with directors permitted to take into account the interests of persons other than shareholders 'to the extent the law allows this' (an issue discussed later). Disclosure requirements should be ample but efficient, so that cost does not outweigh utility. All twelve principles reappear in the Law Commissions' report on directors' duties.²⁵ However, the Law Commissions also produced a refined list of four 'headline' principles that could properly guide the consideration of individual issues. These are accessibility, certainty, graduated regulation of conflicts of interests, and efficient disclosure.²⁶ The Law Commissions suggested that identification of these principles (even

²⁴ Law Commission, *Shareholder Remedies* (1997, Law Com No 246) paras 1.9–1.12.

²⁵ LCR para 3.4.

²⁶ LCR para 3.16.

without weighting them, and even acknowledging that not all were relevant to all issues) would nevertheless permit conceptual questions to be identified, and would provide a benchmark for testing the different views expressed by respondents.²⁷

Against this background, the CLRSG produced its own somewhat different list of four guiding principles.²⁸ First, it advocated a presumption against interventionist legislation, and in favour of facilitating markets, especially by the provision for transparency of information. Secondly, it favoured minimum complexity and maximum accessibility, with its related principle of 'think small first'. Thirdly, it was against creating criminal offences. Finally, it asserted that the enforcement jurisdiction should be allocated to the most suitable regulatory body.

Had all of these aspirations been reflected in the current reform proposals, the final scheme defining directors' duties might have been predicted to be a relatively slim package which collected together rules of a similar character, described the content of the relevant obligations, and designated the most appropriate enforcement mechanism (and perhaps remedial consequences). It is still not clear how the drafting will appear in its final form, but as yet there is no suggestion that related issues will be collected together in this fashion. To take perhaps the simplest example, there are certain transactions to which the directors are not at liberty to commit the company. Many of these arise because of proscriptions in the general law, and clearly it would be inappropriate to list them in any statutory formulation of directors' duties, whatever the resulting benefits of 'accessibility'. But there are other limitations contained in the companies legislation itself. With these, it seems more in line with the adopted guiding principles of reform to collect them together, rather than to leave them littered throughout the hundreds of sections currently comprising the Companies Act 1985. This would mean, for example, that limitations relating to capital maintenance rules, share issues, loan transactions, substantial property transactions and specific remuneration options for directors, and limitations derived from a company's own objects and articles, would all be collected together as transactions barred absolutely, or barred subject to meeting particular enabling provisions, or barred subject to compliance with certain disclosure obligations. These statutory rules could then be linked to rules defining the consequences of breach for the company, for affected third parties, and for the defaulting directors, with specific indications of how (and by whom) those consequences should be enforced. It would seem easy enough to incorporate the edict to 'think small first' within such a streamlined statutory framework, and to concentrate on deleting excessive and unnecessary proscriptions in a company-specific context. Moreover, it would also seem possible – logically necessary, in fact – to address the issue of who is best placed to enforce these proscriptions, and what remedy is best designed to secure compliance. The goal is to demand compliance only where restrictions are deemed essential, and to secure that compliance in the most cost-effective and measured way possible. In part, this goal requires that the rules be easily accessible to the parties affected.

The use of economic analysis

In a novel move, both the Law Commissions and the CLRSG have overtly incorporated economic analysis into the reform project. The Law Commissions'

²⁷ LCR para 3.2.

²⁸ SF v and paras 2.21–2.26.

consultation paper attracted criticism for this, but the report proceeded undeterred, although with some paragraphs devoted to explaining and justifying the stance.²⁹ In the end the criticisms seem unwarranted and the justifications unnecessary. Disappointingly, beyond providing a technical justification for some fairly common-sense propositions, it is difficult to see what the analysis adds to the project. This is notwithstanding the sophisticated specialist advice given to the reform bodies. Again, this result seems to follow from an overly narrow focus at the start of the project. Economic analysis should have been used, if at all, to help define and justify the grand scheme, not to deal with each detail in a relatively *ad hoc* fashion. This is not to deny the value of other theoretical inputs. Both the Law Commissions and the CLRSG concede that 'law and economics' is only one of several potential perspectives that might be used to evaluate the existing law and any alternative proposals. Other policy issues may be equally important, or even more important, in the final analysis. However, none of these other inputs is given serious analysis. Instead, 'economic efficiency' is itself defined so widely that this model alone could be used to advance any one of a huge range of commercial or social objectives. The Law Commissions suggest that outcomes can be judged economically efficient not merely on the traditional Pareto or Kaldor-Hicks tests, but equally where the outcomes deliver technical efficiency,³⁰ allocative efficiency,³¹ or dynamic efficiency.³² Which is to be preferred in any given context is not described. The resulting breadth is such that any programme could surely be judged 'efficient' on one or other of these tests. Some other theory or policy is needed to prioritise the various issues and interests at stake.³³

One possibility, favoured by the Law Commissions, is a theory of risk. There is little sense that economic efficiency underpins their normative classification of exposure to risk, although economic analysis is then used to select the most appropriate – the most efficient – form of rules.³⁴ Using this approach, the Law Commissions suggest that 'mandatory rules' should be reserved for situations where third parties are clearly at risk; that 'strong default rules' (eg requiring shareholder approval for disapplication) should apply where shareholders are at exceptional risk; and (most commonly) that 'weak default rules' (requiring nothing more than meaningful disclosure, subject to confidentiality and cost, for their disapplication) should apply where shareholders are at only limited risk.³⁵

In this limited 'positive' sense, directed at providing a theoretical analysis of the most appropriate enforcement mechanisms, economic analysis comes into its own.³⁶ The CLRSG expressly adopts the relevant learning.³⁷ Even here, however,

29 LCR paras 2.4–2.9.

30 ie 'minimisation of waste' (without defining 'waste'). LCR para 2.11.

31 ie 'allocation of resources to those who are able to make best use of them in the sense of valuing them most highly': LCR para 2.11.

32 ie 'promotion of systems (in this context, companies and industries) which can compete and survive in a rapidly changing environment': LCR para 2.11.

33 Given that the aim is to devise 'a modern company law for a competitive economy', the goal is surely not only an 'efficient allocation of resources', but also the encouragement of options and opportunities for innovation.

34 See S. Deakin and A. Hughes, 'Economic Efficiency and the Proceduralisation of Company Law' [1999] CILR 169.

35 LCR paras 2.16–2.17, 2.28, 2.32, 3.57, 3.72; LCCP para 3.72.

36 A.I. Ogus, 'Economics and Law Reform: Thirty Years of Law Commission Endeavour' (1995) 111 LQR 407, 407, 417, 420.

37 See, eg, CS ch 13 (Sanctions), but not, surprisingly, ch 12 (Regulatory and Institutional Framework). Also see the DTT's website for S. Deakin, 'Economic Effects of Criminal and Civil Sanctions in the Context of Company Law' <www.dti.gov.uk/cld/review.htm>.

there seems to be little more than a recognition that pressure will produce a response – and the greater the pressure the more likely the desired response, since all parties are presumed to behave as rational self-interested actors. More subtlety is needed if this principle is to be balanced against the desire to have minimal regulation and the presumption against criminal sanctions.

Foundation assumptions and the core duty of compliance and loyalty

As already noted, directors' duties are defined and circumscribed by the way in which the company is conceived. The CLRSO saw the alternatives as being to equate the company with the shareholders or, alternatively, with wider, 'pluralist', interests.³⁸ The CLRSO plumed for the 'enlightened shareholder value' approach, seeing this as requiring directors to maximise shareholder value, but to do so in the light of social and ethical constraints – ie in an 'inclusive' manner.³⁹ This might seem like having your cake and eating it too, especially since the CLRSO expressly conceded that its objectives were pluralist: companies should be run in a way which maximises competitiveness, wealth and welfare of all.⁴⁰ This intermediate strategy was seen as most likely to advance competitiveness because it best reflected the needs of companies to meet up with contemporary concepts of value in the modern economy.⁴¹ Accordingly, the CLRSO's proposal is that this core duty of loyalty should appear as a statutory obligation imposed on directors to achieve the success of the company for the benefit of the shareholders by taking proper account of all relevant considerations. The proposed formulation ('Principle 1') is worth setting out in full:⁴²

1 Compliance and loyalty

- a A director must exercise his powers honestly⁴³ and for their proper purpose, and in accordance with the company's constitution and decisions taken lawfully under it [or under the general law⁴⁴].
- b Subject to that requirement, he must (so far as he practically can) exercise his powers in the way he believes in good faith is best calculated in the circumstances, taking into account both the short and the long term consequences of his acts, to promote the success of the company for the benefit of its members as a whole.
- c The circumstances to which he is to have regard for that purpose include, in particular, (as his duties of care and skill may require):
 - aa the company's need to foster its business relationships, including those with its employees and suppliers and the customers for its products and services;
 - bb the impact of its operations on the communities affected and on the environment; and
 - cc its need to maintain a reputation for high standards of business conduct.

38 SF para 5.1.15.

39 SF para 5.1.41; DF paras 2.11–2.17; CS para 3.5. The American Law Institute has adopted a provision which confers discretion on directors to subordinate the interests of shareholders to the wider ethical considerations which may reasonably be regarded as appropriate to the responsible conduct of business: *Principles of Corporate Governance* (1994) para 2.01 b (2).

40 DF para 2.21.

41 DF para 2.3.

42 DF para 3.40; and CS paras 3.11–3.24, confirming that the statement was intended to provide a hierarchy of obligations, beginning with compliance with the constitution.

43 Although this requirement appeared in DF para 3.40, some responses to the consultation suggested that its ambit was uncertain, and the CLRSO now proposes to delete the requirement: CS para 3.13.

44 CS para 3.15 suggests the need to add these further words to include informal decisions which may be taken by shareholders.

This duty of 'compliance and loyalty' seems intended to modify, not emasculate, the traditional duty imposed on directors to act *bona fide* in the interests of the company and for proper purposes.⁴⁵ As reformulated, directors have a primary duty to act within the law and for proper purposes (the duty of compliance) and a secondary duty to act in good faith to promote the success of the company for the benefit of the members as a whole (the duty of loyalty). The duty of compliance is not controversial, although it leaves it to the courts to determine which purposes are 'proper', a task which has proved difficult in the past.⁴⁶ The duty of loyalty deserves further comment.

The CLRSG was persuaded that there were strong policy arguments against adopting the pluralist approach.⁴⁷ In particular, if the directors were given a *power* to decide issues in a 'pluralist' manner, then there would be too much unpoliced discretion residing in the board of directors; on the other hand, if the directors were given a *duty* to decide issues in this fashion, there would be too much unpoliced discretion in the courts. Both results were seen as unacceptable. In any event, the strategy was seen as unnecessary to achieve pluralist objectives. These, it was argued, could be better addressed by specifically directed, independent legislation (eg specific environmental or labour laws).⁴⁸ This view closely mirrors that adopted by the Hampel committee. In its view a pluralist conception would leave directors with no accountability to anyone and no yardstick against which their performance could be measured.⁴⁹ This is no doubt why the CLRSG insists that the overriding interest of the members is central to its concept of the duty of loyalty; its aims cannot, it says, be equally well achieved simply by reference to the success of the company.⁵⁰

Two comments can be made in response to this. The CLRSG suggests that it is merely reformulating the traditional duty so as to encompass wider (more modern) perspectives. This is doubtful. First, it seems that the proposed duty of loyalty is potentially *narrower* than the existing law. At present the directors are obliged to consider the interests of the company's employees.⁵¹ Under the proposed reforms it seems that they will be obliged to consider these interests only to the extent that this will promote the success of the company *for the benefit of its members*. Take a stark example. With this refinement, it is not at all clear whether the famous Henry Ford could have decided, legitimately, to pay his employees more and charge his customers less.⁵² If he could not, then it seems we have returned to 'the bad old days'. If he could, because his proposals could be justified by 'taking account of both the short and the long term consequences of his acts', then the discretion residing in directors to favour different interests is so wide as to be ungovernable – a complaint which defeated straightforward adoption of the 'pluralist' vision of the company.⁵³

45 *Hogg v Cramphorn Ltd* [1967] Ch 254; *Howard Smith Ltd v Ampol Petroleum Ltd* [1974] AC 821 (PC). Historically, 'bona fide' was seen as equivalent to 'genuine': L.S. Sealy, "'Bona Fides" and "Proper Purposes" in Corporate Decisions' (1989) 15 *Monash University Law Review* 265. It might even be possible to see the restriction as demanding only 'proper purposes': see *Peter's American Delicacy Co Ltd v Heath* (1939) 61 CLR 457, 511–512 *per* Dixon J.

46 *Howard Smith Ltd v Ampol Petroleum Ltd* [1974] A.C. 821 (PC).

47 DF para 3.24.

48 DF para 2.12.

49 Hampel Committee Report para 1.7.

50 CS para 3.16.

51 Companies Act 1985 s 309.

52 *Dodge v Ford Motor Co* 170 NW 668 (1919).

53 CS para 3.5.

Secondly, this formulation involves a radical departure from the traditional conception of the duty. The duty is equitable, derived from rules applying to trustees managing trust funds for the benefit ('in the interests') of beneficiaries. In this context, the traditional duty does not oblige trustees to select the optimum possible investment for the benefit of the beneficiaries,⁵⁴ nor does it make them liable for the marginal losses where there is a failure to do so. The trustees are simply liable to restore the trust fund if they intentionally (ie dishonestly) exercise their management discretion *contrary* to the beneficiaries' interests.⁵⁵ In short, the duty – described here as a duty of 'loyalty'⁵⁶ – is proscriptive, not prescriptive.⁵⁷ Furthermore, it defines the process, not the end result. If this duty is translated into the corporate context, then it should prohibit directors from intentionally⁵⁸ exercising their management discretion contrary to the *company's* interests (or contrary to the success of the company). It seems certain that the CLRSG would prefer this to be formulated to address the *members'* interests, but that seems to introduce unnecessary complexity. When directors exercise a discretion to devote the company's assets to stakeholders who are not legally entitled to insist on the additional benefit – be they shareholders,⁵⁹ employees, customers, or the community at large – it seems far simpler to judge the directors' actions as 'loyal' if they are designed to advance the success of the *company* and disloyal if they are not. This can be assessed directly without adding the gloss that the directors' intention ought to be to benefit the members as a whole, whether in the short or long term. With commercial companies, success is measured in money – the relevant aim is to increase the value of the company.⁶⁰ If 'members' are nominated as the primary focus, it will never be clear whether the directors ought to favour members' short or long-term interests, or their interests in income or capital generation. A 'body of shareholders', even a theoretical one, will rarely be sufficiently homogenous to provide clear-cut guidance. Moreover, there is no need to nominate the members as the focus of the duty solely to ensure an appropriate enforcement mechanism. The duty is owed to the company, whatever formulation is adopted.⁶¹ The company has the right to sue, and any compensation recovered from defaulting directors will accrue to the company. The shareholders will be motivated to ensure that breaches are restrained and claims are pursued⁶² because it is they who are entitled to the residual value in the company, not because the duty demands that the directors focus on their consolidated benefit.

54 Whether assessed objectively or subjectively.

55 This is a subjective limitation on exercise of the discretion; the proper purposes doctrine provides an additional objective limitation.

56 More accurately, and more usually, described as a duty of 'good faith'. The duty of 'loyalty' is the fiduciary duty imposed on directors and trustees requiring them to prefer their duty to the company, or to the beneficiaries, over their own personal interests. The CLRSG's use of the familiar terminology of 'loyalty' to mean something quite different is likely to cause confusion.

57 Of course, even when the trustees' discretion is exercised legitimately in the beneficiaries' interests, the trustees will be liable for any losses caused by the negligent exercise of their discretion. However, this sounds in a different duty, the duty of care (see DF para 3.40, Principle 5).

58 Adopting the broader equitable concept of fraud.

59 Shareholders certainly belong in this series: although they have residual rights to the company's assets on a solvent winding up, they are not entitled to demand a particular dividend distribution from a solvent company. This is true even for preferential shareholders. Their entitlement is to preferential payment if a dividend is declared.

60 The fact that the members will benefit is incidental, as is the fact that the stock market – if it is involved – will reflect the increased value of the company as an increase in the price of the shares.

61 DF ch 3 and paras 3.20–3.31.

62 Whether by installing a board of directors prepared to take the necessary action, or by pursuing a derivative action if the circumstances warrant it.

The CLRSG, and many reform bodies before it, have tried to use the law to improve the standards of behaviour of directors. They have turned to the duty 'to act in good faith in the interests of the company' (however 'the company' might be conceived) as the appropriate vehicle to advance this project. This seems misconceived. Historically, the duty was not designed to perform this function. It had an altogether more limited function, as described above. It is one thing for the law to prevent directors acting *against* the interests of the company, or for *improper* purposes, or *without* reasonable care, or *self-interestedly*;⁶³ it is quite another to impose upon them an enforceable duty to act in a manner 'best calculated to ... promote the success of the company'.⁶⁴ The law does not generally attempt to compel good behaviour in this fashion; it merely sets minimum acceptable standards. True, it can *encourage* good behaviour. It does this in many ways. Tax incentives are used to encourage environmentally friendly practices or charitable giving; disclosure regimes (such as the proposed OFR⁶⁵) and internal monitoring systems⁶⁶ encourage adoption of best practices. But these aspirational behaviours are not formulated as binding legal *duties*. This, it is suggested, is the real reason why there has been so much rather fruitless debate on the 'scope' issue. We do indeed want to know the answer to the question posed, but not in order to translate it into a legal duty in manner suggested by the CLRSG. If the proposed statutory duty of 'loyalty' is to have teeth, then it might be advisable to demand less of it, and to ensure that the lesser demands are enforceable.

Other duties imposed on directors

The draft statement of directors' duties⁶⁷ also sets out principles relating to independence of judgement, conflicts of interest, fairness, and care, skill and diligence. The draft might have been expected to attract most controversy in relation to conflicts of interest, but it has not: there appears to be general approval of the suggested rules.⁶⁸ This duty deserves closest attention, but something should also be said of the other duties. Principle 2, describing a duty to exercise independence of judgement, is not controversial. Independence is essential to proper performance of a director's functions, so much so that it is surprising the duty was not incorporated within Principle 1 (discussed in the previous section). A duty of compliance and loyalty surely obliges directors to exercise their powers *independently*, in good faith, for proper purposes and in accordance with the law. As it is, the draft statement does not expressly demand that directors exercise independence, merely that they do not restrict their powers to do so. Any other formulation would probably be impossible to police separately from the existing duty of compliance and loyalty.

63 This is the nub of the fiduciary duty of loyalty – not that the director act in the interests of the company (with the then impossible task of determining whether this has been done), but that where there is a conflict between his *duty* to the company and his own *personal interests*, he should favour his duty to the company. This is an enforceable proscription.

64 DF para 3.40 (Principle 1(b)).

65 The disclosure required by the Combined Code (appended to the Stock Exchange Listing Rules) works in the same way.

66 This, it seems, is the function of the shareholders' power to dismiss directors without reason (Companies Act 1985 s 303). But while the directors are in post, their activities (and the shareholders' rights) are regulated by law, not by shareholder directives: see S Worthington, 'Corporate Governance: Remedying and Ratifying Directors' Breaches' (2000) 116 LQR 638.

67 DF para 3.40.

68 CS paras 3.25–3.29.

Principle 4, describing the duty of fairness, requires directors to 'act fairly as between the company's members'. It is designed to restate the existing law,⁶⁹ which requires directors not to discriminate unreasonably between shareholders.⁷⁰ This formulation also puts in a positive way the incantation underpinning section 459 of the Companies Act 1985, which can provide shareholders with a remedy where the directors (or others) have acted in a manner which is 'unfairly prejudicial'. It may seem that nothing changes where a double negative is replaced by a positive, but on one view a law which gives shareholders the *possibility* of a remedy where the directors have acted in an unfairly prejudicial manner is something different from a law which places directors under a positive duty to act fairly. Articulating the remedy for breach of this duty would go a long way towards clarifying the intended reach of this provision.

Principle 5, imposing on directors an obligation to act with care, skill and diligence, comes as no surprise and appears to have attracted no adverse comment.⁷¹ This duty is simply one manifestation of the general duty of care imposed on all individuals in conducting their activities. If an individual is negligent and thereby causes harm to someone, then the law requires the individual to pay compensation. All of this is straightforward, and there would seem to be no reason to exempt directors from the general operation of the law. However, two specific matters have caused concern in the corporate context for much of the history of this duty.

The first is that the standard of care imposed is commonly perceived as too low to be effective in regulating the conduct of directors. Shareholders suffer, and, if the company fails, creditors suffer too.⁷² The role of directors has changed dramatically in the past 100 years. Then, directors were usually part-time, non-executive members of the board, not expected to devote much of their time to company business and not expected to bring any special skill to bear on their deliberations.⁷³ Judges and commentators generally seemed unable to recognise that the relevant law was capable of adapting to contemporary commercial expectations to demand more of directors and to assess those demands objectively.⁷⁴ The draft statutory statement will side-step this debate and effectively ensure that directors adhere to a minimum, objectively assessed, reasonable standard of care.⁷⁵ Notably, however, the CLRS has not adopted the Law Commissions' recommendation that the statutory duty explicitly recognise not only the function of the director but also the circumstances of the company – ie its size and type.⁷⁶ As well, if directors have even greater skills or knowledge than this minimum requirement, then they will be expected to use those greater skills for the benefit of the company and the standard of care in those circumstances will be assessed subjectively. Section 310 of the Companies Act 1985 will continue to

⁶⁹ DF para 3.65.

⁷⁰ Certainly they cannot use their powers for this *purpose*: that would be in breach of their duty of compliance, which requires them to act for proper purposes. If this duty is intended to mean nothing more than that, then there is no need for a separate statement.

⁷¹ It is not mentioned in CS.

⁷² The company's creditors in fact received better protection, by statutory fiat, some years earlier than the shareholders: see Insolvency Act 1986 s 214.

⁷³ See eg *Re City Equitable Fire Ins Co* [1925] Ch 407.

⁷⁴ Although see *Re D'Jan of London* [1994] 1 BCLC 561 (Hoffmann J); S. Worthington, 'The Duty to Monitor: A Modern View of the Director's Duty of Care' in F. Patfield (ed), *Perspectives on Company Law*: 2 (London: Kluwer, 1997) ch 11, and the references cited there.

⁷⁵ The format will follow that of Insolvency Act 1986 s 214, although with simplified wording.

⁷⁶ LCR para 5.15.

apply, so that companies will be unable to opt out of this requirement. All of this is perfectly acceptable, although it does seem that even this dual standard duty of care is not as onerous as the standards effectively demanded by the Company Directors Disqualification Act 1986.⁷⁷ Alignment seems desirable.

The second difficulty in applying the law of negligence to directors is that it is unusually difficult to say what sort of conduct should count as negligent when directors are carrying out their management functions, at least where their conduct is neither grossly negligent nor fraudulent. Companies are set up to take risks; directors are employed because they are good at this – they are entrepreneurs. The fact that a company makes a loss, or even fails totally, is not of itself indicative that its directors have been negligent in the management of the company's affairs. This led to earlier calls for the introduction of a statutory 'business judgement rule', as operates in the United States and several other countries. Such a rule would confirm that directors who make loss-generating decisions reasonably, after collecting all the appropriate evidence, and after appropriate consultation and evaluation, will not be found guilty of negligence in the management of the company, and so will not be held liable to compensate the company for the losses suffered. The CLRSG has not adopted this approach, rightly it seems.⁷⁸ There appears to be no need for such a rule. The normal rules of negligence clearly admit that damage and harm do not, of themselves, prove negligence; and a director who offered proof of the matters outlined above would clearly not be found negligent, even in jurisdictions without a business judgement rule in place. The same analysis applies where a director delegates particular functions, and loss results.⁷⁹ What is notable, however, is that this commercial assessment is clearly left entirely to the courts.

The draft statutory statement of directors' duties contains one further principle, Principle 3. It states the duty which is to apply in relation to conflicts of interest. Although it seems not to have attracted adverse comment, it does raise concerns. These concerns can best be addressed in the light of the text of the proposed duty:⁸⁰

3 Conflict of duty and interest

3 A director must not:

- a authorise, procure or permit the company to enter into any transaction in which he has an interest unless the interest has been disclosed to the relevant directors to the extent required under the Act;^[81] nor
- b use any property, information or opportunity of the company for his own or anyone else's benefit, nor obtain a benefit in any other way in connection with the exercise of his powers, unless he is allowed to make such use or obtain such benefit by the company's constitution, or the use or benefit has been disclosed to the company in general meeting and the company has consented to it.

This draft text clearly separates two distinct factual situations in which conflicts of duty and interest are likely to arise. Paragraph (a) deals with cases where the director contracts with the company; paragraph (b) with cases where the director

⁷⁷ See eg the discussion in *Re Barings Plc (No 5)* [1999] 1 BCLC 433 (Jonathan Parker J); [2000] 1 BCLC 523 (CA).

⁷⁸ DF paras 3.69–3.71.

⁷⁹ The CLRSG does not propose to deal with the issue explicitly in the statute. DF para 3.71. Similarly, LCR para 5.36. The Law Commission reached this view on the basis that, since the law was still developing, it was impossible to state the position simply and accurately.

⁸⁰ DF para 3.40, Principle 3.

⁸¹ This is a reference to Companies Act 1985 s 317, requiring a director to disclose [material] interests to the board of directors.

contracts (or has some other arrangement) with third parties, not with the company. This division is commonly adopted for pedagogic purposes: the remedies in the first cases may, it seems, be limited to rescission of the underlying contract, with no option to request an account of profits should rescission be unavailable according to prevailing equitable rules; the remedies in the second case are not so limited, with disgorgement of the director's profits being effected either by means of a constructive trust or a personal obligation to account.⁸² Despite these remedial differences, it is clearly recognised that both situations are simply fact patterns raising the possibility of directors being tempted to act in their own personal interests rather than in accordance with their duties to the company. The law seeks to ensure that the company's interests are favoured by adopting a strict [fiduciary⁸³] rule that whenever a director acts in these circumstances, any profits he might make will have to be disgorged to the company. The risk to the company is the same in both cases. The way for the director to avoid this consequence is to obtain the fully informed consent of the person to whom he owes his duty – ie the company. Historically, it has been seen as necessary for the general meeting, not the board of directors, to provide this consent. This may not be defensible.⁸⁴ However, it seems even less defensible to suppose that one fact pattern warrants the protection delivered by a general meeting vote while the other does not. This is even more so when the interested director is to be permitted to vote at the board of directors' meeting.⁸⁵ This will mean that, provided the appropriate disclosure is made, and regardless of the influence exercised by the interested director, it will be impossible to challenge a self-dealing transaction on the grounds of conflict of duty and interest. It may, of course, be possible to mount a challenge based on improper purposes,⁸⁶ absence of good faith, or negligence – but the remedies for those breaches do not include disgorgement. The primary rule needs to be articulated in a more defensible fashion. This could be a rule which requires approval by a (disinterested) general meeting,⁸⁷ or one which requires approval by a disinterested board of directors – but a rule which requires this procedure in *all* cases of conflict of duty and interest.⁸⁸

On the other hand, it may be true that the real agenda being pursued by the CLRS is to re-classify these self-dealing transactions as transactions in which the company cannot participate unless the board of directors has notice of the director's interest. If this is the case, then the breach of duty in issue is not (or not solely) a breach of fiduciary duty. It is, instead, a breach of the duty of compliance, and the remedies should follow accordingly. These remedies would require *all* the

82 See Worthington, n 66 above, and the references cited there. The traditional statement in the text is currently under attack by both academics and the judiciary. The article describes and evaluates the current state of play.

83 The word 'fiduciary' is deliberately omitted from the statutory statement of duties. It is not clear what, if anything, this adds to the statement, especially since the intention is that the judges will have resort to existing 'fiduciary' and 'non-fiduciary' case law: DF para 3.44.

84 See the arguments in Worthington, n 66 above.

85 See DF para 3.62, indicating that the duty will not even arise unless the interested director authorises, procures or permits the transaction.

86 And if the defaulting directors carry the board's decision to approve, then it seems the argument will become circular – the approval will not have been properly given, so the directors will have breached their duty not to profit in circumstances of conflict.

87 And of course the general meeting of a particular company could delegate its power to the board of directors (as is commonly done in a limited context by adoption of Table A art 85), unless this were prohibited by statute. Whichever body voted, however, the vote would have no weight unless it was judged to be an *effective* exercise of voice by the company: see Worthington, n 66 above.

88 *ibid.*

directors involved (knowingly?) in committing the company to the impugned transaction to compensate the company for its losses, but no more than that; they would not require the interested director to disgorge his profits.

The difficulties do not end there. Paragraph (b) of Principle 3 describes what is traditionally termed the 'corporate opportunity doctrine'. It restates the doctrine in an orthodox fashion, without attempting to address the practical difficulties which have dogged application of the equitable rule. In particular, the proposed statement does nothing to assist the courts in resolving the perennially difficult issues of whether an opportunity is 'the company's' or not, especially where the director heard of the opportunity in his private capacity, or where the company declined to, or could not, exploit the opportunity itself, whether for practical or legal reasons. Nor is any consideration given to the quantification of the profit the director will be deemed to have made as a result of the breach, rather than as a result of the exercise of his own skill and energy. In short, the court will be left with the same discretion – and the same difficulties – which mark the present equitable duty. This Principle therefore provides a stark illustration of the magnitude of the input which will be required from the courts in order to define these duties. It also makes it doubtful whether any statement can do more than alert directors to the potential scope of their duties.

The CLRSG considered and rejected the case for a sixth general principle requiring directors to consider the interests of creditors where the company's solvency was threatened. This was seen as unnecessary, being already covered by the inclusive loyalty principle.⁸⁹ Nevertheless, the duty was recognised as sufficiently distinctive to warrant adding a warning on Form 288, which directors need to sign on their appointment. Interestingly, and notwithstanding all this detail, the functions of the *board* of directors are not addressed explicitly. The draft statement refers only to individual directors; the collective duty of the board to supervise and monitor the management of the company is not mentioned, even though this is widely recognised as the key function of this body.⁹⁰

Role of Part X

Part X of the present Companies Act 1985 contains almost 40 sections regulating fair dealing by directors. The provisions are principally concerned with directors' remuneration, substantial property dealing with the company, loans from the company (and related dealings), and share dealings. These provisions far exceed the equivalent regulation in Canada or in other EU member states.⁹¹ Nevertheless, the CLRSG sees retention of Part X as necessary to prevent abuse, even though it concedes the provisions are complex and rather incoherent. The proposed statutory rules discussed in the previous section were seen as clarifying the law, but not as sufficiently strengthening it so as to deal with these issues. In general, the CLRSG saw fit to adopt the Law Commissions' proposals in relation to Part X, and simply annexed those proposals to its own report.⁹² The preferred view is that one or two of these sections be repealed as no

⁸⁹ DF paras 3.72–3.73. Such a duty was also seen as cutting across the rules in Insolvency Act 1986 s 214, which allow action by a liquidator. This conclusion seems to misunderstand the role of the liquidator.

⁹⁰ *Re Barings*, approved by the CA, n 77 above.

⁹¹ LCR para 2.7.

⁹² DF Paras 3.86–3.89 and Annex C; LCR Part 6ff.

longer relevant, but that in the main the provisions should stand,⁹³ although with some substantial drafting amendments to clarify both the duties themselves and the relevant remedies. Here, attention is directed only at section 317 and the remedial provisions.

The Law Commissions devoted considerable attention to section 317 of the Companies Act 1985.⁹⁴ As it currently stands, this section imposes a fine on directors who fail to disclose to the board of directors any interests they have in transactions with which the company is involved. It is proposed that the section be clarified to indicate that the necessary disclosure need only be of material interests of which the director is (or ought to be⁹⁵) aware and which are not already known to the board, and that disclosure need only be to the directors present at the relevant board meeting.⁹⁶ Again, as the law currently stands, breach of the section does not render the transaction voidable, although that is the separate consequence of the application of non-statutory fiduciary rules. The Law Commissions recommended that criminal penalties be removed and civil penalties be introduced. The CLRSG has deferred its decision on criminal sanctions, but considers that civil penalties ought to be imposed. In addition, the CLRSG proposes that the section provide for the transaction to be voidable if the interest is not disclosed, and that the director be liable to account for profits and compensate for losses. It further proposes that this civil liability be mandatory, regardless of the terms of the company's constitution. The CLRSG saw all of this as restating the common law.⁹⁷ This is doubtful. As already noted, it is not even clear whether liability for self-dealing transactions extends beyond rescission to include an account of profits when rescission is unavailable;⁹⁸ certainly directors in breach of their fiduciary duties of loyalty are not liable (under this head⁹⁹) for losses suffered by the company. This loss-recovery response would be more appropriate if the section were directed at corporate capacity, or at directors' authority, rather than at fiduciary duties of loyalty designed to generate disinterested decision-making. As noted earlier, the CLRSG is also of the view that, provided disclosure is made, interested directors should not be prohibited from voting on the self-dealing transaction.¹⁰⁰ This too seems wrong.¹⁰¹ At present the matter is regulated by companies' articles, although the articles of listed companies *must* prohibit interested directors from voting on matters in which they have a material interest (subject to limited exceptions). In Australia, too, directors of public companies are prohibited from voting in these circumstances.¹⁰² This section (which, in its revised form, seems to be repeated in the proposed general statutory provisions dealing with directors' duties¹⁰³) seems

93 This means that a new Companies Act will still contain provisions dealing with payments in connection with loss of office (ss 312–316), declaration of interest (s 317), disclosure of, and period for, service contracts (ss 318–319), substantial property transactions (ss 320–322), and credit transactions (ss 330–344).

94 LCR Part 8.

95 LCR para 8.31: there should be no obligation to disclose where the director does not know of the interest and could not reasonably have been aware of it. The same sort of disclosure is required of the interests of connected persons.

96 DF Appendix C paras 11–13.

97 DF Appendix C para 18.

98 See the earlier discussion of Principle 3.

99 They may, of course, be concurrently liable for other breaches which have as their remedy the obligation to compensate for losses.

100 DF Appendix C para 17. LCR paras 8.89–8.91.

101 See the earlier discussion of Principle 3.

102 Australian Corporations Law s 232A, although this prohibition may be lifted by the board.

103 DF para 3.40, Principle 3(a).

overly lenient in its statement of the duty, and may therefore prove inadequate in dealing with the practical problems it seeks to address, and yet it seems excessive in the remedies it imposes for breach.

On the remedies front, and again adopting the Law Commissions' view, the CLRSg proposes a comprehensive code of civil remedies to deal with all the provisions in Part X. The argument for this is that it will allow comparisons to confirm consistency and remove anomalies, and will only require one lot of litigation to determine the meaning in several contexts.¹⁰⁴ As far as the particular remedies are concerned, the Law Commissions' proposals largely follow the model currently familiar in the context of sections 320–322, although there are some important deviations:¹⁰⁵ self-interested transactions will be voidable subject to the usual bars, including when the court so directs;¹⁰⁶ whether the transaction is avoided or not, the director concerned will be liable to account for profits and /or compensate for losses, again subject to any limitations or restrictions on quantum imposed by the court in their discretion.¹⁰⁷ The particular suggestions about the differential applications of this discretion are not yet clear.¹⁰⁸ The potential impact on defaulting directors has already been seen in cases such as *Re Duckwari plc (No 2)*,¹⁰⁹ where the defaulting director was held liable for the losses suffered by the company due to falls in the property market. The only concession seems to be that directors will not be held liable for losses that the company would have suffered in any event.¹¹⁰

These ideas (bar the discretion to be given to the courts) have been around for a long time and seem to have caused few real problems, and yet they seem to confuse two quite distinct regulatory impulses. A company law statute might wish to bar certain types of transactions, or bar them unless they meet certain hurdles. This decision can be made for any number of valid reasons, and may be designed to protect any one of a wide range of stakeholders. In these circumstances it seems entirely appropriate to look to the directors who committed the company to a transaction in breach of such restrictions and require them to cover the losses thereby incurred. Alternatively, the regulation may be designed to prevent directors using their wide discretionary powers, or even their positions as directors, to line their own pockets rather than to promote the company's business. Here there is no defined end goal which directors must reach. The issue is not entry into prohibited ventures, or negligent assessment of corporate commitments; it is the drive to obtain disinterested service from directors. The appropriate remedy is disgorgement of any profits a director makes from acting self-interestedly. It may be that the situations dealt with in Part X ought to be classified under both heads. It seems, however, that a specific statutory rule is needed only in relation to the first type of regulation; the second type will be covered automatically by the

¹⁰⁴ LCR paras 15.1–15.7

¹⁰⁵ LCR para 8.97ff.

¹⁰⁶ LCR para 8.113.

¹⁰⁷ LCR paras 8.113–8.115, 8.119, the last paragraph suggesting that the court's discretion should also be able to modify the duty to account.

¹⁰⁸ LCR para 15.5. The suggestion is that so far as s 322 is concerned, the discretion to limit liability should not be exercised even if the company suffered no prejudice; however, so far as s 317 is concerned, the court should exercise a discretion, and it should be a relevant consideration that the company suffered no loss. The logic of this is not apparent, unless it relates to the 'seriousness' of the breach. Both transactions are self-dealing transactions: one is a substantial property transaction requiring shareholder approval, the other an interested transaction requiring disclosure to the board of directors.

¹⁰⁹ [1998] 2 BCLC 315 (CA).

¹¹⁰ LCR para 8.117.

more general conflicts provisions.¹¹¹ There is no problem with this form of concurrent liability, although such separation would effectively modify the breadth of the remedies currently contemplated by Part X. In addition, there seems to be no real reason to give the court a particular statutory discretion in these cases when the general provision in section 727 of the Companies Act 1985 appears apt to cover the concerns raised. In fact, it might be thought that these very concerns which militate in favour of a strong judicial discretion also serve to suggest that the remedial provisions are drafted too widely.

In short, it seems that the better way to deal with these Part X provisions is to deal with them in the section of the Act which is devoted to corporate capacity (including the powers of the various organs), not in the section on directors' duties.¹¹² This section could readily accommodate a complete list of restrictions on corporate activity and an indication of who is limited by them and who is liable under them. This separate treatment would seem to fit better with the economic efficiency arguments presented in all the reform documents and discussed earlier.

Statutory statement of duties

One of the central issues for corporate law reform has been whether matters would be improved with a statutory statement of directors' duties. Over the course of the review process, each alternative has at some stage enjoyed preferred status. The Law Commissions' consultation document suggested that a statutory code would be too restrictive, inhibiting the essentially dynamic and evolving nature of the law on directors' duties.¹¹³ It preferred, instead, to have non-binding, educative, authoritative pamphlets outlining the principal duties owed by directors. After consultation, however, the Law Commissions' report advocated partial (ie non-exhaustive) codification of the 'settled' duties, leaving the courts to develop any unsettled or new duties.¹¹⁴ The advantages of this approach were seen to lie in its flexibility and potential dynamism.

The CLRSG advocates a third option. It proposes an exhaustive¹¹⁵ statutory statement of directors' general duties,¹¹⁶ drafted at a sufficiently high level of generality to overcome the Law Commissions' criticisms of statutory codes. The statement is intended to combine flexibility with accuracy and accessibility, and to embrace the CLRSG's views on 'scope', thereby giving comprehensible and useful guidance to company directors and others.¹¹⁷ This does appear to be the best

111 Although then the inadequacies of Principle 3(a) become readily apparent: see the earlier discussion.

112 See, too, the earlier discussion of Principle 1 (duty of compliance) and Principle 3(a).

113 LCR para 4.27, although the Law Commission was also of the view that some duties were not still 'evolving' but were now 'settled' – LCR para 4.28 – so that a statutory formulation of them would improve accessibility. This approach, and its detailed articulation, present some surprises. The Law Commission suggested that a director's duty to act bona fide and in the interests of the company was a paradigm example of a settled duty which could be codified; by way of contrast, a director's duty not to misuse corporate property was seen as still too unsettled to be appropriately subjected to codification (LCR paras 4.15–4.18). Many might have thought that quite the reverse was true, especially in the light of the vehement and ongoing 'stakeholder' debate, which is all about the proper remit of the 'interests of the company'.

114 LCR paras 4.28–4.44. It continued to recommend non-binding pamphlets for information: LCR paras 4.64–4.67.

115 DF paras 3.15, 3.82. The statement is not intended to be exhaustive outside its stated field: see the similar arguments in *Re Wait* [1927] 1 Ch 606 in relation to the then current Sale of Goods Act.

116 DF paras 3.37–3.73.

117 By contrast, the Law Commission was of the view that a general statutory statement could not be informative: LCR paras 4.15–4.18.

option. The arguments in favour of the alternatives are not persuasive. The suggestion is also in line with several other Commonwealth jurisdictions.¹¹⁸ However, the justification for this approach is surely only to improve the clarity and accessibility of the law. The CLRSG's aim is not (for the most part) to alter the law;¹¹⁹ nor, it seems, will the statement greatly affect the law's development. It is unlikely that the old cases will lose their force as illustrations of the proper application of the law. Indeed, given the generality of the principles (a generality deliberately designed to enhance flexibility), it is unlikely that the statement will do anything to impede developments and refinements in the law which might otherwise have taken place at common law.¹²⁰

If the primary function of the statutory statement is to make the law clearer and more accessible, then there are some potential flaws in the proposed regime. The heading to the proposed statutory statement, 'The general duties of a director of a company', is intended to reinforce the idea that there are other specific duties imposed on directors of companies.¹²¹ This is counter-productive. It seems preferable to collect together (at least) the Companies Act provisions which impact on the duties of directors, whether the provisions affect the existence or the exercise of their authority.

In addition, two other matters require further attention. First, the Law Commissions' and CLRSG's papers suggest the provisions are to apply only to directors formally appointed to the board. There is no discussion of other parties involved in corporate governance who might equally well be subjected to the same regime.¹²² Unless the net is cast wider, two classes of management will evolve: directors subjected to the Act, and other officers subjected to the common law and equitable rules, especially the rules relating to fiduciaries. Secondly, there is no detailed discussion of the extent to which these duties can be ameliorated or avoided. The issues are controversial. It seems better to address the problems in the context of each specific duty, rather than trying to re-draft the Companies Act 1985 s 310 to provide a comprehensive general rule.¹²³ Finally, and notably, although the DTI recognised that governance can be imposed by primary or secondary legislation, codes of practice, Stock Exchange Listing Rules and such like,¹²⁴ all the documents to date concentrate exclusively on the first route.

Statutory statement of civil remedies for directors' breaches

The proposal for a statutory enactment of directors' duties is being combined with a move to enact civil remedies for their breach. Space prevents a full discussion of the issues here. The CLRSG invites responses on the general desirability of codification, and asks for specific comments on the approach advanced in a paper prepared by Mr Richard Nolan and published on the DTI website.¹²⁵

118 Australia, Canada and New Zealand – SF ch 4. Canada also has a statutory scheme of remedies, and Australia has a system of civil penalties.

119 DF para 3.43.

120 This is despite the CLRSG's conviction to the contrary: DF para 3.15 n 22.

121 DF para 3.43.

122 Contrast the width of the statutory provisions in Australia, where the law is designed for a widely defined class of corporate 'officers': Australian Corporations Law s 232(1).

123 Of course, a general provision may suffice to deal with the important issue of D&O insurance.

124 MCL para 5.8.

125 CS paras 13.71–13.74, and R. Nolan, 'Enacting Civil Remedies in Company Law' <www.dti.gov.uk/cld/review.htm>.

As with the statutory statement of directors' duties, a statutory enactment of remedies seems to be an achievable goal, and one worth working towards. Again, as with the duties themselves, the aim should be to make the law clearer and more accessible, not to reform it. To that end, it seems that all that is needed is a simple indication of the party liable, and the consequences of liability – be it rescission of the underlying transaction, compensation for losses incurred by the company, or disgorgement of the profits generated by the breach. The general law would continue to provide refinements to the quantification exercise undertaken by the courts. In this way, the statute would not need to specify rules relating to remoteness of damages, contributory negligence, general defences, tracing consequences, and so forth. In essence this is what is proposed in respect of the statement of duties, and if the duties are to be allowed to develop in the usual common law, incremental fashion, then so too must the remedies, otherwise application of the law is likely to become distorted.

Although the general idea of a statutory statement of remedies seems warranted, the particular form proposed in the paper adopted by the CLRS is not in line with the arguments advanced here. The paper affirms the importance of aligning specific remedies with specific duties.¹²⁶ Nevertheless, it then adopts what is largely a 'Part X-style' regime, so that most breaches of duty (bar the duty of care and skill) attract a 'menu' of remedies – ie rescission, liability for losses and gains – available against a wide class of defendants. The problems associated with this approach have already been discussed.¹²⁷ It effectively reallocates risk in a potentially unwarranted and unacceptable fashion. The various duties are fundamentally different in their objectives, and the traditional remedies by and large seek to achieve the law's aims in the least interventionist manner possible.¹²⁸ The statutory statement of remedies will presumably incorporate the proposed revised section 727 of the Companies Act 1985, which gives the court a discretion to excuse directors who have behaved honestly and ought fairly to be excused.¹²⁹ It is then a matter of policy whether the statute should also describe the rules which permit a company, rather than the court, to excuse its defaulting directors. The drive to provide clear and accessible rules suggests a positive answer, but the technical issues are difficult.¹³⁰

Enforcement mechanisms

Only two points need to be made in relation to the enforcement mechanisms proposed for directors' duties. So far as civil enforcement is concerned, the reform documents confirm that directors' duties are owed to the company and are to be enjoyed collectively. Whether or not a breach should be pursued depends upon the collective interests of the members, not their individual interests. It follows that it is for the board to decide whether or not to sue.¹³¹ This is orthodoxy, but it needs to be carefully integrated with the newly proposed duties of compliance and loyalty (Principle 1) and fairness (Principle 4).

¹²⁶ *ibid* para 18.

¹²⁷ See the earlier discussion of Part X.

¹²⁸ See S. Worthington, 'Fiduciaries: When is Self-Denial Obligatory?' [1999] CLJ 500; Worthington, n 66 above.

¹²⁹ DF paras 3.76–3.77.

¹³⁰ DF paras 3.79–3.81, 4.120–4.126. See Worthington, n 66 above, for one assessment of the rules.

¹³¹ DF para 3.79.

So far as criminal enforcement is concerned, there is an unexpected preference for the retention of criminal sanctions for a wide variety of wrongs. This is unexpected because the DTI views the current legislative regime as over-burdened with unnecessary criminal provisions where civil sanctions would be preferable. Indeed, one of the specific reform principles adopted by the DTI was a principle against creating criminal offences; another was to allocate the enforcement jurisdiction to the most suitable regulatory body.¹³² This view persisted in the *Strategic Framework Document*, where it was predicted that rationalization of the framework would result in fewer statutory prescriptive rules with criminal sanctions.¹³³ The Law Commissions, too, had suggested that criminal sanctions ought to be retained only where there would otherwise be problems in achieving necessary compliance, or where there were cost disincentives which mitigated against shareholder litigation.¹³⁴ This conclusion was in line with one of the Law Commissions' guiding principles, the 'appropriate sanctions principle', which indicated that there should be a range of effective and realistic sanctions backed by appropriate oversight and enforcement machinery.¹³⁵

Basic economic analysis might suggest that criminal sanctions are only warranted where the social cost of non-compliance is high, and where cost-benefit analysis suggests that enforcement by a non-state party is unlikely. The two parts of the equation are in balance: a very high social cost of non-compliance might suggest that criminal sanctions are appropriate even if alternative enforcement mechanisms are available and likely to be utilised; as the social cost goes down, however, the argument for criminal sanctions can only be sustained if that appears to be the only available enforceable mechanism. It goes without saying that a rule need not, and should not, be imposed unless its enforcement (by whatever mechanism) is regarded as desirable. It follows that the need to impose criminal sanctions often depends upon how effective civil sanctions are.¹³⁶ Certainly, some directors' duties are sufficiently important that public enforcement is warranted (the duties of disclosure, for example), but the current proposals suggest an over-dependency on criminal sanctions which appears out of line with the reform principles adopted at the outset.¹³⁷

Conclusions

The stated aim of the DTI's current review is to produce a simple, rational framework which is modern and competitive, and which facilitates enterprise and promotes transparency and fair dealing.¹³⁸ It is against that ambitious goal that its efforts will be assessed. As yet the rhetoric seems not to be matched by the substance, at least in the area of directors' duties. Nevertheless, there is much that can be built on. The criticisms articulated in this note can, in substance, be boiled down to an assertion that the framework must build unambiguously on coherent general principles. The core ideas must be refined and particularised to reveal practical rules. Much will rest in the drafting. The framework structure will be as important as the particular provisions. The devil is not entirely in the detail.

132 SF v.

133 SF para 5.5.5.

134 LCR paras 2.35–2.36.

135 LCR para 3.4.

136 LCR para 2.18.

137 This is notwithstanding the analysis provided in CS ch 13.

138 MCL Foreword and para 3.8.

Integrated Ombudsmanry: Joined-up to a Point

*Brian Thompson**

Following an approach in October 1998 by the Parliamentary Commissioner for Administration (PCA), the Health Service Commissioner (HSC) and the Commissioners for Local Administration in England (CLA), the Cabinet Office conducted a review of the public sector ombudsmen in England which reported in April 2000¹ and a consultation paper based on the review was issued in June 2000.²

The key conclusions of the *Review* were:

- the current legislative provision for the ombudsmen is restrictive and is distorting the service and therefore requires a radical overhaul;
- proliferation of methods of public service delivery will continue, which will mean finding new ways to help complainants and enabling the ombudsmen to work with others to provide an integrated service;
- reshaping of the ombudsmen is needed to respond to reshaped government³ and
- the MP filter can no longer be sustained in an era of joined up government and it is strongly recommended that it is abolished.⁴

This analysis led to the fundamental recommendation that, instead of three separate ombudsmen systems focused on central and local government and the National Health Service (NHS), these should be combined in one single commission for England; and access for complaints on central government matters (to the PCA) should no longer be through a MP (the MP filter) but direct as is the case with the CLA and HSC. The consultation paper focuses in particular upon those recommendations and others on powers, jurisdiction, appointment and accountability. In addition, consultees were invited to comment on any of the review's recommendations or any aspect of the organisation and operation of the public sector ombudsmen in England.

The approach which the *Review* takes is one of concentrating on the resolution of complaints. Its analysis of the legislation and the working practices of the ombudsmen lists various barriers. The PCA has frequently been criticised for the length of time taken to complete an investigation and action has been taken to reduce the backlog of cases.⁵ The PCA was the first ombudsman in the United Kingdom (UK) and served as a template for its counterparts, first in Northern Ireland, then the NHS, local government and most recently in the devolved

* Liverpool Law School.

1 P. Colclutt and M. Hourihan, *Review of the Public Sector Ombudsmen in England* (London: Cabinet Office, 2000). It is also available on the Cabinet Office Website at <<http://www.cabinet-office.gov.uk/central/2000/ombudsmenreview.pdf>>. The review was announced in the White Paper *Modernising Government*, Cm 4310 (1999).

2 *Review of the Public Sector Ombudsmen in England: A Consultation Paper* (London: Cabinet Office, 2000).

3 n 1 above, para 2.43.

4 *ibid* para 3.52.

5 In the Annual Report for 1999–2000 the PCA stated that the backlog had been eliminated, HC 593 (2000) para 1.

arrangements for Scotland and Wales. The analysis of the working practices of the PCA characterised them as an audit type investigation. The conception of the PCA was that of an assistant to MPs in their historic role of holding the executive to account and seeking the redress of citizens' grievances. The investigation would end in a report to the MP. The complainant was not a direct customer of the PCA. The focus of the investigation was maladministration, and maladministration which had produced injustice. Fairness and thoroughness had to be built into the investigation process given the possibility of a report which concluded that the investigated body had caused injustice through maladministration. A combination of the legislation and its implementation by the first PCA, Sir Edmund Compton a former Comptroller and Auditor General, led to an investigation process which, in effect, required three separate inquiry stages. First, the PCA could only accept complaints for investigation which were within jurisdiction and so they had to be screened;⁶ secondly, the principal officer of the body complained against was given an opportunity to respond to the complaint⁷ and then the PCA investigation would take place.⁸ The *Review* does not pick up on a factor which can delay the issue of the report in this third investigation stage, which is the practice of showing a draft report to the principal officer.⁹ This allows facts to be checked and a reaction to a draft recommendation on a remedy where maladministration has been found to cause injustice. The reaction can, on occasion, turn into a protracted negotiation.

These stages contribute to a longer timescale than would be the case if the focus was simply to resolve a complaint. The CLA took the initiative in seeking to resolve some complaints informally, that is without beginning or completing an investigation, and this is now something which the PCA does.¹⁰ As the legislative basis for such informal resolution is unclear, the *Review* recommends that it be authorised in the statute¹¹ which will create the new commission for the public sector ombudsmen in England.

The *Review* recommends that the ombudsmen adopt a new approach as follows:

- a focus on outcomes and in particular on complaints resolutions;
- a clear recognition of when intervention is required;
- initially an informal approach aiming to achieve co-operation and perhaps using a conciliatory approach;
- investigation in the old sense to be used if informal methods do not work.¹²

Within this framework instead of the segmented process described above, the process would be continuous, moving away from 'acceptance or rejection' of a complaint to the action or resolution proposed (in some cases, 'no action' by the ombudsman). Thus a complaint would be pursued until it is clear that the

6 Parliamentary Commissioner for Administration Act 1967, ss 4–6, Schedules 2 & 3.

7 *ibid* s 7(1).

8 *ibid* s 1(1).

9 The practice of showing a draft only to the investigated body was found not to be in breach of natural justice in *R v Parliamentary Commissioner for Administration, ex parte Dyer* [1994] 1 WLR 621. However, the *Review* recommends, at para 6.20, that a copy of the draft also be shown to the complainant, which has been the practice of the CLA.

10 In the Annual Reports for 1998–99 HC 572 (1999) and 1999–2000, HC 593 (2000) the PCA records that 90 and 121 complaints respectively were resolved without an investigation. The figures in the CLA Annual Reports for local settlements were 2,414 with no report issued, 24 with an issued report in 1997–98; and 2,251 without a report, 42 with a report in 1998–99.

11 n 1 above, para 6.63.

12 *ibid* para 6.61.

complaint is out of jurisdiction and not investigable, a settlement to the ombudsman's satisfaction has been achieved, or investigation has been completed and a report has been produced.¹³ In the 1999–2000 Annual Report, issued after the publication of the *Review*, the PCA talks about extending the 'product range' so as to encompass informal resolution, investigations of varying intensity and of the amalgamation of the screening and investigation directorates from April 2000.¹⁴

The *Review* devotes some space to the recommendation to abolish the MP filter¹⁵ mainly, one assumes, because a majority of MPs wish to retain it, whilst most commentators favour abolition. A survey was conducted for the *Review* in which support for retention, at 52 per cent, had declined from the 1993 level of 58 per cent.¹⁶ The *Review's* survey indicated that Labour MPs were generally in favour of abolition whilst Conservatives were not and that MPs first elected in 1997 were more in favour of abolition than those with a greater length of service. The survey responses also had MPs who admitted uncertainty about jurisdiction and general access arrangements. As complaints must be referred by MPs then the fact that around 70 per cent of complaints submitted do not end in informal resolution or an investigation report shows that MPs are not absolutely clear about the criteria. To be fair some MPs may knowingly refer complaints outwith the statutory criteria but this is suspected to be a small number. In arguing for the removal of MPs as gatekeepers to the PCA, it is pointed out that this does not mean removal of MPs entirely from the field of complaints about constituents' grievances with central government. Some will begin the complaint with MPs or move on to them if unhappy with the reaction of the respondent body. Since the days of the Whyatt report¹⁷ which initially proposed the creation of an ombudsman in the UK, the number of grievance-handling agencies has increased and now ranges from bodies' own complaints machinery whether internal or independent such as the Independent Cases Examiner for the Child Support Agency, to the ombudsmen and various regulatory regimes for the privatised utilities. Given this complexity, for citizens and MPs, it is best if there is a single gateway to the public sector ombudsmen which can exercise a more consistent policy on access to the ombudsmen. This first contact is rightly regarded as very important because some complainants, having been told that they should take their complaint to another body, decide to drop it. The *Review* recommends that there should be monitoring of complaints which do return after having been directed to another grievance-handler and that steps should be taken to find out what happens to those which do not return.¹⁸ Where a complainant is directed to another body, it should be subject to the condition that if it has not been resolved within a specified time-limit, then it will be automatically referred to the commission.¹⁹ There is concern that the gateway should not go beyond signposting and being an impartial investigator. Accordingly the recommendation is made that by entering into partnerships with the Community Legal Service and the Community Health

13 *Ibid* para 6.54.

14 HC 593 (2000) paras 1.15–1.20.

15 n 1 above, paras 3.33–3.56.

16 Survey carried out for the Select Committee on the Parliamentary Commissioner for Administration and reported in *Powers, Work and Jurisdiction of the Ombudsman*, HC 33 (1993).

17 Justice, *The Citizen and the Administration: The Redress of Grievances* (chaired by Sir J Whyatt) (London: Stevens, 1961).

18 n 1 above, para 6.43.

19 *Ibid* para 6.35.

Councils, those bodies will gain access to information which can help them support complainants.²⁰

The proposed organisational structure²¹ does not do much more than amalgamate the three ombudsmen. All of the ombudsmen would continue to be appointed by the Queen, they would report to Parliament and be subject to scrutiny by the National Audit Office. The new commission would be chaired by one of the ombudsmen for the purposes of representing it externally, for management purposes and when there was a requirement to answer to Parliament. It is proposed that this ombudsman would be responsible for matters relating to the UK as a whole and for reserved matters in Scotland, Wales and Northern Ireland. The internal organisation would operate as a single entity for the management of work and generally for accountability, policy-making, funding and resource management. The individual ombudsmen would be appointed as office-holders with a personal jurisdiction across the entire work of the new commission. They should not be appointed to have particular functional or geographical responsibilities. However by agreement within the new commission they would each be identified with a particular group of the bodies under its jurisdiction. Thus, for example, local authorities would know which member of the new commission would deal with them individually or corporately on questions of policy and practice. It was recommended that the staff of the new commission should specialise in aspects of the functions of bodies under jurisdiction and as necessary form teams to deal with partnership working by those bodies. Such partnerships may involve bodies not under jurisdiction, or under the jurisdiction of another complaints investigation scheme, and innovative collaborative arrangements would be needed. Each ombudsman would be responsible for his or her own cases and not subordinate to any other ombudsman. No ombudsman would be superior to another in making decisions and recommendations about matters under jurisdiction nor would any ombudsman act in any appellate capacity if a complainant disagreed with another ombudsman's decision.

It is also proposed that the commission have (a) non-executive board members who could be drawn from 'external bodies or the general public' and whose role would be to assist in the monitoring of the impact of the new commission and to offer operational and policy advice;²² and (b) associates, including the public sector ombudsmen in Scotland, Wales and Northern Ireland, and the Data Protection/Information Commissioner.²³ Currently the PCA is an *ex officio* member of the Commission for Local Administration in Wales which can exchange cases for investigation with its English counterpart, usually on the basis of convenience or propriety.

The reasons for proposing the unified commission are first that the delivery of public services crosses the boundaries of the separate public sector ombudsmen remits and this is likely to increase,²⁴ and second to try to increase awareness of and access to the ombudsmen and indeed, other grievance-handlers. As the *Review*

20 *ibid* para 6.31. The Legal Services Commission's Funding Code criterion 5.4.7 states that requests for legal representation and support funding may be refused if there are alternatives to litigation, including ombudsman schemes, which should be tried first: <http://www.legalservices.gov.uk/stat/fc_criteria.htm>.

21 n 1 above para 5.3.

22 *ibid*.

23 n 1 above, paras 7.3–7.4.

24 This is 'joined-up' government, a prominent feature of the White Paper, n 1 above.

does not make any recommendations on changes of jurisdiction²⁵ we have the situation that whilst it is proposed to remove the MP filter for PCA cases, where bodies are subject to the PCA and have their own internal and possibly independent complaints examiners, then these must usually be resorted to before the PCA can be involved. In the NHS recourse to its two-stage complaints scheme is also normally expected before the HSC can become involved. CLA normally expects a complainant to have raised the matter with the local authority, if not to have exhausted its complaints system, otherwise the complaint will be regarded as premature and will not be accepted.

Continued role confusion?

The response which the House of Commons Public Administration Select Committee (PAC) made to the consultation paper on the *Review* was both positive and critical.²⁶ It agreed with the idea of a single commission,²⁷ and the removal of the MP filter²⁸ but preferred that legislation should specify those bodies which were not within the ombudsmen's jurisdiction rather than those who were.²⁹ The committee did raise concerns about the composition and jurisdiction of the proposed commission. It suggested that a review of the ombudsmen should be accompanied by a review of their jurisdiction and referred to the evidence of academic witnesses who argued first for the inclusion of contractual issues and personnel matters and second for the consideration of the merits of issues where there was no maladministration.³⁰ The committee also thought that some areas of the public sector were not included and suggested that they should be, giving education as an example.³¹ They were concerned about the number of associate ombudsmen and the non-executive members.³² The Select Committee noted that the Prison Ombudsmen was not included in the new commission.³³ In the *Review*, the reason given was that this was a 'niche' role, not established by statute and properly part of the executive.³⁴ The Prison Ombudsman, like the Independent Cases Examiner at the Child Support Agency, is above the departmental complaints system but under the PCA. Both are meant to be independent whilst funded by the departments they oversee.

One of the problems with the *Review* is that it seems prepared to put up with the complicated position on access conditions and would seek to improve it by requiring the new commission to agree protocols with other 'intermediary' complaint handlers, on arrangements to deal with cases, which should be publicised,³⁵ thus

25 It is suggested that the new legislation might adopt a variety of methods on specifying which bodies are within jurisdiction. Listing generic types such as local authorities and NHS Trusts would be sufficient to cover many bodies within jurisdiction; it might, however, be necessary to list at least some organisations to banish all doubt, n 1 above, para 5.8.

26 HC 612 (2000).

27 *ibid* para 7.

28 *ibid* para 12. The Select Committee noted that originally the MP filter was meant to assuage MPs' feelings but felt that it was now inconsistent in a world of public service charters not to have direct access to the public sector ombudsmen.

29 The consultation paper, n 2 above, asked if the legislation should specify those outside jurisdiction.

30 Respectively raised by Dr. P. Giddings and Professor P. Birkinshaw, n 26 above, paras 13, 16.

31 *ibid* para 14.

32 *ibid* para 8.

33 *ibid* raised specifically by Professor M. Seneviratne.

34 n 1 above, para 4.7.

35 *ibid* para 7.6.

helping complainants and respondent bodies. This could be regarded as a pragmatic approach but it could also be seen to reflect continued confusion over the real role of the ombudsmen. For some academic commentators there has long been a view that ombudsmen would do better to concentrate on audit and the improvement of performance rather than resolving complaints or as Harlow and Rawlings have put it on fire-watching rather than fire-fighting.³⁶ The PCA was conceived as a last resort so that complaints which had alternative remedies available would be outside jurisdiction unless the PCA determined that it would be unreasonable for the complainant to seek that other remedy.³⁷ The trend of thinking and practice on complaints handling reflected in the Citizen's Charter and its successor Service First³⁸ and the NHS complaints scheme,³⁹ is that complaints should be dealt with speedily and at, or close to the point of contact with the complainant. There is a pyramid structure so that complaints not resolved at that lower level are dealt with by independent review and at the top of the structure is the ombudsman. The Service First and NHS guidance stresses the importance of learning from mistakes so as to enhance performance.

The *Review's* recommendations stress resolution and speed and urge that meetings and phone calls be used in preference to paper-intensive methods. The call to produce searchable electronic databases of decisions which can be used by complainants and their advisers as well as by respondent bodies⁴⁰ seeks to balance fire-fighting with fire-watching as does the proposal that informal resolution should not foreclose the possibility of a full investigation which could bring to light systemic problems.⁴¹ Concern for resolution prompts the recommendations that where the unified gateway to the commission has to refer the complainant to another grievance-handler, the situation should be monitored to see if the complainant returns, and also to set conditions so that the complaint would be automatically transferred. One aspect which could assist the audit role is the ombudsmen's ability to conduct investigations on their own initiative without the trigger of a complaint. Noting that ombudsmen in other countries have this power, it is said that it is not used much and that it could imperil the ombudsmen's impartiality and so it is specifically rejected.⁴² It would be unfortunate if bodies subjected to an own initiative investigation doubted the impartiality of the ombudsmen given that random audit is not uncommon. This is a useful tool and has been conditionally advocated by select committees.⁴³

36 C Harlow and R. Rawlings, *Law and Administration* (London: Butterworths, 2nd ed 1997) 427–432.

37 Parliamentary Commissioner for Administration Act 1967, s 5(2). A similar provision applies to the CLA, Local Government Act 1974, s 26(6).

38 Citizen's Charter Complaints Task-Force, *Effective Complaints Systems Principles and Checklist* (London: Cabinet Office, 1993); *Putting Things Right* (London: Cabinet Office, 1995); *Good Practice Guide* (London: Cabinet Office, 1995), *Service First, How To Deal With Complaints* (London: Cabinet Office, 1998).

39 Department of Health, *Being Heard: Report of a Review Committee on NHS Complaints Procedures* (chaired by Professor A. Wilson) (London: Department of Health, 1994); NHS Executive, *Complaints: Listening ... Acting ... Improving ... Guidance on Implementation of the NHS Complaints Procedure* (NHS Executive, 1996).

40 n 1 above, para 6.40.

41 *ibid* paras 6.4, 6.9.

42 *ibid* para 6.15.

43 Successive Select Committees on the PCA have suggested initiation of investigations but on different conditions: it would serve as a follow-up to reports where maladministration and injustice had been found, HC 615 (1978), para. 31; again where there had been a report finding maladministration and injustice and the PCA suspected that a citizen had suffered but not made a complaint, the PCA could ask the constituency MP to make a reference, HC 129 (1990), para 22, and where a select committee made a reference to the PCA, HC 33, (1994) para 44. The Public Administration Committee in its

The bigger partiality problem in complaints systems is the perception of complainants that only independent grievance-handlers are appropriate. They fear that where a member of the organisation complained against is conducting the investigation, then the loyalty felt amongst colleagues means that the complaint will not be dealt impartially. This has bedevilled the system for police complaints and the new NHS complaints procedure,⁴⁴ and led to the rejection of proposals which would have left complaints against local government to their own internal complaints systems with no investigatory role for the CLA.⁴⁵

The *Review* rejects calls made for a Public Administration Commission with a wider oversight and standard setting remit, emphasising that the ombudsmen's focus must be on complaints resolution.⁴⁶ There is awareness of the total complaints picture and a recommendation that the current focal point in the Cabinet Office be slightly expanded to serve as the body to monitor and co-ordinate public sector complaints processes.⁴⁷ Welcome as the integrationist approach of the *Review* is, it does not go far enough. There is a need for a body which not only has an overview of the whole field of complaints handling but can use the complaints it processes to inform suggestions for improvement. The cultural change represented by Citizen's Charter/Service First in which complaints are accepted and learnt from, rather than evaded and denied, still has some way to go. Aspects of the New Public Management in which quantitative measures are meant to be driving toward improvement can have unintended consequences so that, for example, complaints may not be recorded if there are league tables which display complaints received. If complete complaints recording does not happen then it prejudices monitoring and learning from complaints.

The concern in the *Review* that nothing should detract from a concentration upon resolution does not rule out a commission containing all aspects of complaints handling: resolution, learning from mistakes, research and dissemination of good practice. Logic might suggest that not only should one learn from the example of the Audit Commission with its financial irregularity and value for money audits and the promotion of best practice, but that a combined Public Administration and Accounts Commission would be the route to follow. This might be too much but it should be noted that there is a recommendation that whilst care should be taken in determining whether the investigation focus might widen to take systemic points into account in case the original complaint is lost sight of, there should be good liaison with the Audit Commission and the National Audit Office in order to take forward matters of concern.⁴⁸

response to the consultation upon the *Review* also thought that there was a role for own initiative investigations, in certain circumstances, which were not specified, and also at the suggestion of the committee itself, HC 612 (2000) para 11.

44 See the Home Office, *Complaints Against the Police: A Consultation Paper* (London: Home Office, 2000) and the critique of, and reform proposals for, the independent review stage in NHS complaints, H. Wallace and L. Mulcahy, *Cause for Complaint? An Evaluation of the NHS Complaints Procedure* (London: Public Law Project, 1999).

45 The reaction to the proposal to lose independent investigation of complaints against local authorities in Sir G. Chipperfield, *Report of the Financial Management and Policy Review of the Commission for Local Administration in England Stage I* (London: Department of the Environment, 1996) led to its abandonment and the confirmation of the investigatory role of the CLA in A. Whetnall, *Report of the Financial Management and Policy Review of the Commission for Local Administration in England Stage II* (London: Department of the Environment, 1996).

46 n 1 above, para 6.86.

47 *ibid* para 5.19.

48 *ibid* para. 6.89.

Academics have advocated the need for research and it is a function and capacity that an overarching body in the field of administrative justice ought to have.⁴⁹ One important topic is the appropriate number of levels in the complaints pyramid. The factors which have to be balanced are speed, cost, and fairness which involves impartiality and independence. Is there a need for three levels given the time the entire process takes, or do the different tiers deal with appropriate cases, and are there variations in the different areas of the public sector? The proposed commission would be in an ideal position to undertake such research and given the recommendation to have non-executive members on its board, this suggestion does not push very far beyond the *Review's* boundaries. Indeed this would seem to be the only sensible role which the proposed non-executive members could play. As the PAC commented, their role is 'undeveloped' in the *Review*.⁵⁰

The *Review* does take into account the relationship with the courts. Referring to Lord Woolf's 1996 *Access To Justice Final Report* they declined to follow two of three particular recommendations on this relationship. Ombudsmen's reports should continue not to be binding and not to be enforced by the courts, nor was there support for the idea of the courts being able to refer a case before them to ombudsmen, but it could be useful for ombudsmen to be able to refer an issue of law arising in an investigation to the courts for determination.⁵¹

Conclusion

The *Review* has one big idea, amalgamation, but this is not followed through comprehensively. There are many good recommendations which could improve awareness of, and access to, the ombudsmen and the lower tiers in the public sector complaints system. The proposals on working methods are unlikely to offset the rise in complaints which abolition of the MP filter would bring about.⁵² While outside the *Review's* remit,⁵³ the logic of a merged commission also applies to Scotland and Wales, and in Northern Ireland where it could formalise the 'virtual' situation in which there are two ombudsmen systems, both offices always held by one person working in shared accommodation. This would, however, tip the

49 See, for example, B. Thompson, 'Administrative Justice: Towards the Millennium, Towards Integration?', D. Lewis, 'Filling in the Gaps: A Standing Administrative Conference for the United Kingdom', and M. Partington, 'Future Development' in M. Harris and M. Partington (eds), *Administrative Justice in the 21st Century* (Oxford: Hart, 1999) 463–481, 519–531 and 532–548 respectively.

50 n 26 above, para 8. The committee agreed with the views of the current PCA, Mr M. Buckley, that these non-executive members should not be involved in the handling of individual cases, para 21.

51 n 1 above, paras 7.25–7.27.

52 The Public Administration Committee agreed with Dr P. Giddings that it 'was optimistic, not to say naive' to think no additional resources would be required after abolishing the MP filter, n 26 above, para 12.

53 The Public Administration Committee noted that there was very little discussion of the expected impact of devolution and relations with ombudsmen for Wales and Scotland. As they point out the current PCA holds the offices of Scottish Parliamentary Commissioner for Administration and Welsh Administration Ombudsmen, which eases liaison but this may not always be so, n 26 above, para 24. From October 2000 until January 2001, the Scottish Executive invited comments on *Modernising the Complaints System: Consultation on Public Sector Ombudsmen in Scotland*, SE/2000/84 (2000), which is also available on the consultations website at <<http://www.scotland.gov.uk/consultations/misc/mcas-00.asp>>. This proposed a one-stop shop which would deal with complaints about the Scottish Executive, local government and the NHS in Scotland. It also expressed the preference to remove the 'MSP filter' in respect of complaints against the Scottish Executive, SI 1999 No 1351, art 103(3)(a).

balance between fire-fighting and fire-watching in favour of the former. A Commission for Public Administration with a research capacity could resolve complaints and make suggestions for improving service which a Cabinet Office unit could implement across the public sector.

In the 1970s Australia produced its 'New Administrative Law' following a trinity of reports,⁵⁴ which in turn owed something to the Franks Report. In 2000–2001 there are in addition to this study of ombudsmen, others on the arrangements for judicial review⁵⁵ and tribunals.⁵⁶ This provides an opportunity to attempt a more holistic view of complaints against government and the various redress mechanisms. If it is ensured that judicial review, tribunals, ombudsmen and complaints systems can pay particular attention to the citizen's perspective then this trinity of reviews might produce the 'New Administrative Justice'.

54 *Report of the Commonwealth Administrative Review Committee*, chaired by Sir John Kerr, (Canberra: Commonwealth Government Printer, 1971); *Interim and Final Reports of the Committee on Administrative Discretions* chaired by Sir Henry Bland (Canberra: AGPS, 1973), and *Report of the Committee on Prerogative Writ Procedures* chaired by R. J. Elliott QC (Canberra: AGPS, 1973), see M. Partington, 'The Reform of Public Law in Britain: Theoretical Problems and Practical Considerations' in P. McAulian and J. McEldowney (eds), *Law, Legitimacy and the Constitution* (London: Sweet & Maxwell, 1985) 191 which acknowledged a debt to J. Goldring, 'The Foundations of the "New Administrative Law" in Australia' (1981) XL *Australian Journal of Public Administration* 79.

55 Lord Chancellor's Department, *Review of the Crown Office List*, chaired by Sir Jeffrey Bowman, (London: Lord Chancellor's Department, 2000). The recommendation to rename the Crown Office List the Administrative Court has been accepted, see Practice Direction 20 July 2000.

56 Announced by the Lord Chancellor on 17 May 2000 and to be chaired by Sir Andrew Leggatt. Details and the *Review of Tribunals: Consultation Paper* are available at <<http://www.tribunals-review.org.uk>>

CASES

What Price Auctions Without Reserve?

*Frank Meisel**

The decision of the Court of Appeal in *Barry v Davis (t/a Heathcote Ball (Commercial Auctions) & Co) and Others*¹ brings clarity to an important area, which has been left uncertain for nearly one hundred and fifty years.

It is trite law in auction sales that the advertising of a sale by auction constitutes only an invitation to treat, that a bid constitutes the contractual offer, and that a contract for sale of a lot is effected on the fall of the hammer or other customary means which constitutes acceptance of the offer. It is also the case that any bid that is exceeded by a higher bid which is taken by the auctioneer thereupon lapses and that, in accordance with general contractual principles, a bidder may withdraw a bid at any time before acceptance. All of this may give way to specific provision in the auction conditions.

It has proved difficult to fit into that structure the legal significance of an auction sale which is expressly made 'without reserve'. If a lot is so advertised and the auctioneer refuses to sell to the highest bona fide bidder, either by withdrawing the lot after the bidding has commenced or by accepting a bid on behalf of the seller, he prevents a contract of sale of the goods or land in question arising. Does this render the disappointed 'buyer' without remedy against the auctioneer or seller? The answer given by the majority of the judges in the nineteenth-century case of *Warlow v Harrison*² (albeit *obiter*) was that the unsuccessful buyer could recover against the auctioneer for breach of a collateral contract to sell without reserve. Martin B said³

it seems to us that the highest bona fide bidder at an auction may sue the auctioneer as upon a contract that the sale shall be without reserve. We think the auctioneer who puts up the property for sale upon such a condition pledges himself that the sale shall be without reserve; or, in other words, contracts that it shall be so; and that this contract is made with the highest bona fide bidder, and, in case of a breach of it, that he has a cause of action against the auctioneer.

Until *Barry v Davis* there was, however, no authoritative ruling on the issue and the problem of quantification of damages for breach of such a contract similarly had remained unresolved. The facts were as follows. The defendant auctioneers were instructed to sell two new Alan Smart engine analysers by Customs and Excise who had acquired them in lieu of the manufacturer's VAT liability. The sellers instructed the auctioneers to sell them without reserve and they were so put up for sale. The auctioneer who took the sale, Mr Cross, tried to start the bidding at £5000. There was no bid. He descended to £3000. Still there was none. Eventually the claimant bid £200 for each machine. Mr Cross decided that he could not let

* Senior Lecturer, University of Birmingham. The advice and assistance of my colleagues George Applebey, Richard Goldberg and John Miller is gratefully acknowledged

1 [2000] 1 WLR 1962.

2 (1859) 1 E & E 309.

3 *ibid* 316-317.

them go at that price and withdrew them from the sale. His evidence was that he thought he was 'justified in not selling at auction without reserve if [he thought he] could get more in some other way later.'⁴ Subsequently the machines achieved £750 each in a private treaty sale after advertisement in a magazine.

The county court judge found in favour of the claimant expressly following *Warlow v Harrison* and awarded damages of £27,600 being the difference between the value of the machines and the price at which the claimant ought to have acquired them. The defendants appealed arguing, first, that the auctioneer was entitled to withdraw the lots; second, that there was no collateral contract with the highest bona fide bidder because no consideration moved therefrom; and third, that if any such collateral contract came about it was made through the agency of the auctioneer with his disclosed principal, the sellers, who alone were bound. The appeal, therefore, raised a point (*per* Sir Murray Stuart-Smith) of some general interest and importance.

Entitlement to withdraw lots from sale

The defendant's submission was that the auctioneer makes no express promise to sell to the highest bidder at whatever price merely by stating that the sale is without reserve; no intention to be bound can be found. Putting up a lot amounts merely to an invitation to make an offer to buy and that offer is to be made by bidding. It was also suggested that there might be many situations where it would be legitimate to withdraw lots including where the auctioneer suspects that a 'ring' is operating and where the auctioneer suspects that his seller lacks title.

In dismissing this ground of appeal the Court of Appeal was content to rest on the view of the majority in *Warlow v Harrison* which, although strictly *obiter*, was 'entitled to very great respect.'⁵ An auctioneer who sells without reserve is not permitted to withdraw a lot merely because the highest price offered is not high enough. In so holding, the Court of Appeal distinguished or explained the two cases which appear to have cast doubt on the correctness of the *Warlow v Harrison* dicta: *Mainprice v Westley*⁶ and the Scottish case of *Fenwick v Macdonald, Fraser & Co. Ltd.*⁷

As to whether there might exist situations where the auctioneer was entitled to withdraw a lot, the court was silent but it is submitted that if the highest bidder is a participant in an criminal conspiracy to depress the price he can hardly be regarded as a bona fide bidder and that would be an answer to his claim. The situation where there are doubts about the seller's title is more problematic. The auctioneer gives no absolute undertaking as to his client's title but warrants only that he knows of no defect therein (*Benton v Campbell Parker and Co*⁸) and, in the unlikely event that he discovers the defect only once the bidding on a lot has commenced, a timely announcement should ensure that the highest bidder withdraws before the lot is knocked down. Well-drawn auctions terms of consignment and conditions of sale can provide expressly for rights of withdrawal of lots such as that in the instant case giving the auctioneer the right to re-auction a lot where a dispute arises as to bids.⁹

4 [2000] 1 WLR 1962, 1964.

5 *Ibid* 1966.

6 (1865) 6 B & S 420.

7 (1904) 6 F. (Ct of Sess) 850.

8 [1925] 2 KB 410.

9 On the right to re-offer the lot when there are disputed bids see F. Meisel, 'A Nod and a Wink: The Problem of Disappointed Buyers at Auction' in F. Meisel and P.J. Cook (eds.), *Property and Protection: Essays in Honour of Brian Harvey* (Oxford: Hart Publishing, 2000).

The absence of consideration

This issue was aired by academic writers nearly fifty years ago. Slade argued¹⁰ that making a bid could not constitute consideration for the promise to sell to the highest bidder any more than could saying 'I accept'. On the other hand, in the same volume, Gower suggested that bidders supply consideration by doing an act (bidding) at the request of the offeror.¹¹ Today, both Chitty and Benjamin opine that the consideration is comprised of the benefit to the auctioneer in that the bid provides the potential for driving up the price and detriment to the bidder in that he might become bound. Those views were approved by the Court of Appeal.

The objection may be raised that such a bidder may withdraw his bid at any time before acceptance. How can it be that a revocable offer can create any benefit to the offeree or detriment in the offeror? It was perhaps this aspect that led Lord Wilberforce in *New Zealand Shipping Co. Ltd v Satterthwaite*¹² to put sales at auction at the head of the list of commercial situations where it was difficult to fit the facts without force into the marked slots of offer, acceptance and consideration. It is submitted, however, that a satisfactory answer is provided by remembering that the offer contained in the designation of a lot as being sold without reserve is a unilateral offer to those comprising potential bidders; it is however an offer only capable of being accepted by whomsoever satisfies the condition of becoming the highest bona fide bidder; *ex hypothesi*, such a person will not have revoked his bid. This also solves Gower's quandary that there must be a contract made with each bidder. And the highest bona fide bidder will have supplied consideration by performing the act of making the highest bid.

The analysis is similar to that in the tender cases. In *Blackpool and Fylde Aero Club v Blackpool Borough Council*¹³ the Council invited tenders for the operation of pleasure flights from their airport. The invitations to tender stipulated a time-scale for submitting tenders but the club's tender, despite complying with the requirements, was wrongly treated as being out of time. The club sued for breach of contract in that it was claimed that the Council had warranted that timeous tenders would be duly considered. The Council argued that the invitation for tenders was no more than a proclamation of willingness to receive offers. This submission was rejected. Lord Justice Bingham held:

Where [the] invitation prescribes a clear, orderly and familiar procedure the invitee is in my judgment protected at least to this extent: if he submits a conforming tender before the deadline he is entitled, not as a matter of mere expectation but of contractual right, to be sure that his tender will be considered ...

Thus whilst the tender was unsuccessful, the tenderer could sue on a collateral warranty that complying tenders would be considered.¹⁴ The submission of the tender is both an offer and acceptance of the offer to consider compliant tenders. So, too, in the closely analogous auction field, the bid, whilst a contractual offer to buy, is at the same time also an acceptance of the offer to comply with the rules comprised in the invitation to treat.

10 (1952) 68 LQR 238, 241.

11 *ibid* 457-460.

12 [1975] AC 154 at 157.

13 [1990] 1 WLR 25.

14 *The Blackpool and Fylde Case* was distinguished in *Fairclough Building Ltd v Borough Council of Port Talbot* (1993) 62 Build LR 82 where it was held that an invitor of tenders may decline to consider a tender if there are circumstances warranting it, such as, as here, the existence of a potential conflict of interest and duty.

The immunity of the auctioneer as agent of a disclosed principal

The defendant argued that any collateral contract was made with the principal-seller and not with the agent-auctioneer. This small point was given short shrift:

there is no doubt that when the sale is concluded, the contract is between the purchaser and the vendor and not the auctioneer. Even if the identity of the vendor is not disclosed, it is clear that the auctioneer is selling as agent. It is true that there was no such contract between the vendor and purchaser. But that does not prevent a collateral agreement existing between the auctioneer and bidder. A common example of this is an action for breach of warranty of authority, which arises on a collateral contract.¹⁵

This is certainly the right analysis where, as here, the auctioneer has taken it upon himself to abort the sale in defiance of his seller's clear instructions to sell without reserve. However, in *Warlow v Harrison*, it was the seller who bought in his own goods to prevent a sale at a perceived undervalue. Martin B 'entertained no doubt' that the auctioneer was entitled to an indemnity from his principal in respect of any liability arising from the revocation of his authority.¹⁶

Quantum

Slade¹⁷ suggested that, at most, a disappointed bidder could claim the loss of a chance¹⁸ since, when the goods are withdrawn, there may be others willing to bid higher than the claimant. However, it seems that in the *Heathcote Ball* case the auctioneer withdrew the machines when he had exhausted the bidding with the claimant's bid. (There is a suggestion that the auctioneer had had an offer of £400 from his son-in-law but it is not clear whether this was an auction bid or a private treaty offer.) The Court of Appeal, somewhat oddly, applied section 51(3) of the Sale of Goods Act 1979 which provides for the failure of a seller to deliver goods to a buyer; the buyer gets damages based on the difference between the contract price and the market price of the goods at the date of breach. But there was here no contract of sale between the seller and buyer. Rather there was a contract to effect a contract of sale between the seller and the highest bidder.

The orthodox approach is to regard the object of contractual damages as being to place the claimant, so far as money can do it, in the same situation as if the contract had been performed. If the auctioneer had performed his contract to effect a contract of sale of the machines at the highest price bid (without reserve) the claimant would have acquired machines worth £28,000 (this catalogue price was the assumed value – the position would have been different if the defendant had sought to buy these machines to trade as opposed to use). The price bid was £400. Thus his recoverable loss was £27,600.

It may be thought that the claimant has received a windfall but that is not really so. He has merely recovered for loss of his bargain, both in the technical-legal and in the literal senses of that expression. If a seller at auction wishes to avoid selling at an undervalue he may make the sale subject to reserve but then there is no

15 [2000] 1 WLR 1962, 1967–1968.

16 (1859) 1 E & E 309 at p 317.

17 (1953) 69 LQR 21, 22.

18 In *Thomas Eggar Verrall Bowles (a firm) v Anthony Burnett Rice* (Ch.D Rimer J 21 December 1999) a prospective vendor was wrongfully denied a chance to bid at auction. The court awarded him damages based on the assessment that he had lost a 70 per cent chance of successfully bidding for a lot. He was therefore held entitled to recover 70 per cent of the aggregate of the cost price of the stock plus the likely profit less the cost of acquisition.

guarantee that a sale will be achieved. Thus, a seller who decides to sell without reserve does so to guarantee that a sale will be achieved at a particular time and place, provided that there is one person ready to bid for the property. The downside is that he risks selling at a 'knock-down' price.

Advertising lots as being sold without reserve encourages bidders since they know that provided they beat the competition they will secure the lot; they are not exposed to the possibility that they will successfully outbid everyone else but fail to secure the lot because there is a reserve placed upon it. Provided that the seller is properly advised, he should be able to make a sensible commercial decision whether or no to place a reserve on his auctioned property. But there can be no doubt now in the minds of sellers or auctioneers that if the lot is sold without reserve, they must let the lot go at any price.

One Immunity Has Gone ... Another ... : *Holland v Lampen-Wolfe*

*Mizushima Tomonori**

'Any immunity from suit is a derogation from a person's fundamental right of access to the court which has to be justified.'¹ What was once justified can be held unjustified. In *Arthur J S Hall & Co v Simons*, the House of Lords declared that the advocate was no longer immune from suit.² On the same day, however, another equally long-established immunity from suit was confirmed by the House of Lords in *Holland v Lampen-Wolfe*.³ The immunity at issue was that enjoyed by foreign states. The purpose of this note is to examine a number of issues raised by the House of Lords' decision in *Holland v Lampen-Wolfe*.

The facts of this case are not complex. In arranging general education for those permitted on a US military base in the UK, the US entered into a contract with a US university. Under the contract, the university supplies the personnel for the provision of education. The plaintiff, as part of her employment with the university, gave courses to US military personnel and their families at the military base. Receiving complaints from her students about her conduct, an employee of the US Department of Defense, in his capacity as educational services officer at the base, wrote and faxed a memorandum to the university's office in Germany. He questioned the professional competence of the plaintiff and requested the assignment of another instructor in her place. The plaintiff then brought proceedings for defamation against this officer.⁴ Both the plaintiff and the defendant are US citizens.

* MPhil/PhD candidate, London School of Economics and Political Science. The writer wishes to thank the anonymous referee for helpful comments. The usual disclaimers apply.

1 n 2 below, 581 *per* Lord Hope.

2 [2000] 3 WLR 543.

3 [2000] 1 WLR 1573.

4 In this case, the named defendant was not a foreign state but an individual official. For an analysis of the individual's entitlement to state immunity, see Mizushima Tomonori, 'The Individual as Beneficiary of State Immunity: Problems of the Attribution of *Ultra Vires* Conduct' (2001) 29 *Denver Journal of International Law and Policy* (forthcoming).

A number of questions require consideration: the immunity to be enjoyed by visiting armed forces; state immunity in tort claims as distinct from contract claims; state immunity in defamation claims as distinct from other tort claims. The analysis of these problems is then followed by that of the question why state immunity is compatible with the right to the court.

Section 16(2) of the State Immunity Act 1978 and the applicable law

The foreign state immunity principle had been regulated solely by the common law until the State Immunity Act 1978 came into force. Since then, this Act has provided a statutory basis for the immunity. Section 1(1) of the Act provides: 'A State is immune from the jurisdiction of the courts of the United Kingdom except as provided in the following provisions [sections 2 to 11] of this Part [Part I] of this Act.'

In accordance with section 16(2), however, '[Part I] of this Act does not apply to proceedings relating to anything done by or in relation to the armed forces of a State while present in the United Kingdom and, in particular, has effect subject to the Visiting Forces Act 1952.' The Law Lords found that *Holland v Lampen-Wolfe* related to anything done by or in relation to the US visiting armed forces. This finding, which will not be discussed in this note, disapplied Part I of the Act to this case, and gave rise to the question what then, was the applicable law.

It was suggested by Lord Millett that section 16(2) might have been intended to disapply only the exceptions to immunity, ie sections 2 to 11, rather than Part I as a whole.⁵ Then, section 1(1) remains and immunity would be granted. Yet, this construction clearly flies in the face of the unambiguous wording of section 16(2), and was not to be adopted. Another possible construction, which disapplies immunity altogether, was also rejected, because '[i]t seems unlikely that Parliament contemplated that proceedings might be brought against members of visiting forces in circumstances in which it would not be possible to bring them against state officials who were not members of a visiting force.'⁶ Lord Millett then concluded: '[I]t may be that Parliament assumed that any proceedings falling within section 16(2) would necessarily be covered by immunity at common law.'⁷

The Law Lords thus reached the same conclusion as the Court of Appeal did in this case⁸ and in *Littrell v United States of America (No 2)*.⁹ However, a number of considerations make us doubt whether Parliament intended to apply state immunity at common law in section 16(2) cases.

Section 16(2) is evidently based upon Article 31 of the European Convention on State Immunity 1972, which provides: 'Nothing in this Convention shall affect any immunities or privileges enjoyed by a Contracting State in respect of anything

⁵ Lord Millett says: 'It may merely be the result of mistakenly employing the drafting technique adopted by article 31 of the [European] Convention [on State Immunity 1972], but whereas the 1972 Convention operated to restrict state immunity ("a contracting state cannot claim immunity"), section 1 of the State Immunity Act 1978 gives positive statutory force to the doctrine of state immunity' *ibid* 1585. It seems that Art 15 of the Convention, which corresponds with s 1 of the Act, was overlooked. See also Sir Ian Sinclair, 'The Law of Sovereign Immunity. Recent Developments' (1980) 167 *Collected Courses of The Hague Academy of International Law* 113, 257.

⁶ n 3 above, 1585.

⁷ *Ibid*.

⁸ [1999] 1 WLR 188.

⁹ [1995] 1 WLR 82.

done or omitted to be done by, or in relation to, its armed forces when on the territory of another Contracting State.' In reading this provision, we ought to remember that the visiting armed force has been considered to constitute a special category of state organs for the purposes of immunity, as is the case of diplomats.¹⁰ It would then be more likely that the drafters of this Convention had in mind *visiting armed force* immunity. Indeed, a court of another jurisdiction held: 'The terms of Article 31 recognise that, as a matter of international law, immunities and privileges in respect of anything done or omitted to be done by or in relation to armed forces, when on the territory of another Contracting State, exist.'¹¹

There is certainly some force in the proposition that '[t]he view, formerly widely held, that the [visiting] force was in all respects to be regarded as beyond the jurisdiction of the territorial state (subject to the possibility of a waiver of that immunity) ... can no longer be maintained.'¹² Yet, no evidence is known to the writer which suggests that the drafters of the Convention and those of the State Immunity Act, in 1970s, shared this proposition. Further, applying state immunity at common law to visiting armed forces would amount to viewing them as being on the same footing as other foreign state organs.¹³ Why then is Part I of the Act applied to other foreign state organs but disapplied to visiting armed forces?

Finally, it is worth noting that during parliamentary debates on the State Immunity Bill an amendment to Clause 17(2), which eventually became section 16(2), was moved and agreed to. This amendment is to substitute 'anything done by or in relation to the armed forces of a State while present in the United Kingdom' for 'the armed forces of a State'. Lord McCluskey, in moving this amendment, said:

Clause 17(2) provides that the Bill is not to apply to proceedings relating to the armed forces of a State. It is intended to relate to the activities of forces stationed in the United Kingdom and their supplies, but it could be read to cover also any purchases made here for a foreign State's forces elsewhere. That is not what is intended or is required by [Article 31 of the European Convention on State Immunity].

The Amendment will remove only things done by or in relation to Armed Forces in the United Kingdom from the operation of the Bill, so that purchases made by a State for the use of its armed forces elsewhere (whether of supplies or military equipment) will fall under Part I and enjoy no immunity. For instance, subsection 3(a) of the new clause inserted after Clause 2 will apply to such transactions. ...¹⁴

This amendment hardly makes sense if the applicable law in section 16(2) cases is state immunity at common law, rather than visiting armed force immunity. For, as will be seen, both state immunity at common law and section 3 (ie the new clause inserted after Clause 2) of the State Immunity Act are considered to depend ultimately upon the same test, ie whether the activity at issue is a sovereign activity or not.

10 eg Hans Kelsen, *General Theory of Law and State* (New York: Russell & Russell, 1961) 234, Sompong Suchantikul, 'Immunity of Foreign States before National Authorities' (1976) 149 *Collected Courses of The Hague Academy of International Law* 87, 108; Hazel Fox, 'States in the Market Place' (1994) 110 LQR 199, 203.

11 *McElhenny v Williams* [1995] 3 IR 382, 402 per Hamilton CJ for the Supreme Court.

12 Sir Robert Jennings and Sir Arthur Watts (eds), *Oppenheim's International Law*, vol 1 (London: Longman, 9th ed, 1992) 1157.

13 *ibid* 1157-1158.

14 HL Deb vol 389 col 1533 16 March 1978.

State immunity at common law and tort claims

The Law Lords' conclusion on the applicable law, which is, as has been seen, open to question, made it necessary to ascertain what is the common law rule of state immunity. For this purpose, the Law Lords referred to the *I Congreso* case,¹⁵ which was decided in accordance with the common law as the relevant facts took place before the State Immunity Act came into force, and which brought 'authoritative acknowledgment that English common law has now adopted a restrictive doctrine of state immunity.'¹⁶ The holding of the *I Congreso* case is, in a nutshell, that state immunity attaches only to acts *jure imperii* (sovereign acts) as opposed to acts *jure gestionis* (private law acts).

It is true that, in deciding whether or not a foreign state is immune from suit, numerous courts in various jurisdictions have referred to this classification of state activities. Yet, it is also true that this dichotomy has been established mainly through contract cases. It is therefore not without reason that the question, which is quite relevant to *Holland v Lampen-Wolfe*, was posed: 'Should the general rules on immunity [based upon the above classification] apply to torts as well as to contracts?'¹⁷

It is to be conceded that, as a matter of authority, the *I Congreso* case covers not only contract claims but also tort claims. However, there may be some room for reconsideration. Given the 'accidental' nature of tort,¹⁸ whether damage is caused by an act *jure imperii* or by an act *jure gestionis* is entirely fortuitous and cannot be anticipated by potential victims. Therefore, from the standpoint of equality among victims, and if other considerations can be disregarded, a better rule would be that state immunity is either granted or denied in tort proceedings irrespective of the classification of the allegedly tortious act. It is worth remembering that section 5 of the State Immunity Act, the so-called tort exception, 'is not restricted to tortious conduct performed in the course of commercial transactions and includes personal injuries incurred in the course of public acts *de jure imperii*.'¹⁹

Be that as it may, in so far as the above principle is considered to be applicable in *Holland v Lampen-Wolfe*, the next question is what is an act *jure imperii* or an act *jure gestionis*, that is, how we classify an act into either of these categories. Apparently accepting Lord Wilberforce's proposition in the *I Congreso* case that 'in considering, under the "restrictive" theory whether state immunity should be granted or not, the court must consider the whole context in which the claim against the state is made',²⁰ the Law Lords laid stress on the whole or overall context.²¹ As a result of their consideration of the whole context, including 'the place where the [educational] programme was being provided and ... the persons by whom it was being provided and who it was designed to benefit',²² it was unanimously concluded that the memorandum in question was an act *jure imperii* and attracted state immunity at common law.

15 [1983] AC 244.

16 Hazel Fox, 'State Immunity: The House of Lords' Decision in *I Congreso del Partido*' (1982) 98 LQR 94, 94.

17 Rosalyn Higgins, *Problems and Process: International Law and How We Use it* (Oxford: Clarendon Press, 1994) 85.

18 eg W.V.H. Rogers (ed), *Winfield and Jolowicz on Tort* (London: Sweet & Maxwell, 15th ed, 1998) 20.

19 H. Fox, 'State Responsibility and Tort Proceedings against a Foreign State in Municipal Courts' (1989) 20 *Netherlands YIL* 3, 22. See also Fox, n 10 above, 203.

20 n 15 above, 267.

21 n 3 above, 1577 *per* Lord Hope; 1580 *per* Lord Clyde; 1585 *per* Lord Millett.

22 *ibid* 1577 *per* Lord Hope.

In addition to the ambiguity of the concept of 'whole context', it is not to be overlooked that, consciously or unconsciously, the Law Lords did not refer to another proposition which Lord Wilberforce made in the same case: '[A] private act [ie an act *jure gestionis*] mean[s] ... an act of a private law character such as a private citizen might have entered into'.²³ Indeed, what a commentator emphasises as accepted in the *I Congreso* case is not so much the 'whole context' test as the 'private citizen' test.²⁴

In this respect, it is interesting to note what a Canadian judge stated:

The [private citizen] test was developed in the 'trading cases' and is a useful analytical tool in that context; ... However, I am reluctant to extend the test beyond this class of cases. In particular, I fail to see its utility in the present case [ie union certification proceedings in respect of Canadian civilian staff at a US naval base].²⁵

What was instead adopted by this judge, in order to decide whether the activities at issue are commercial or not, is the 'entire context' approach.²⁶

It would follow that an act which is qualified, in accordance with one test, *jure imperii* in one sort of proceedings can be qualified, in accordance with the other, *jure gestionis* in another sort of proceedings. In other words, one and the same state activity can be, for the purposes of state immunity, an act *jure imperii* and an act *jure gestionis* at the same time.²⁷ Then, one wonders whether any useful purpose is served by suggesting that acts *jure imperii* attract state immunity at common law and acts *jure gestionis* do not.²⁸ Moreover, the question remains unsettled which of those tests (others might be added) is to be applied in which case. It seems that the House of Lords' decision provides little if any guidance for subsequent cases in which state immunity at common law is applied.²⁹

State Immunity Act 1978 and defamation claims

While *Holland v Lampen-Wolfe* was decided as a matter of state immunity at common law, problems of defamation claims will arise also in the context of state immunity under statute. Thus, it would be worthwhile to see how the State Immunity Act deals with these problems. As already mentioned, under this Act, a foreign state is immune from suit unless a specific exception to immunity is applicable.

Before the analysis of the exception invoked by the plaintiff, reference should be made to section 5 of the Act, which concerns some sorts of tort claims. It provides:

23 n 15 above, 262. Equally, Lord Wilberforce quotes with approval (*ibid* 269) the test suggested by Goff J: '[I]t is not just that the purpose or motive of the act is to serve the purposes of the state, but that the act is of its own character a governmental act, as opposed to an act which any private citizen can perform.' *I Congreso del Partido* [1978] QB 500, 528.

Although Lord Clyde mentioned 'commercial transactions such as might be undertaken by private individuals' (n 3 above, 1579), this consideration had no bearing on his reasoning.

24 See Sir Robert Jennings, *The Place of the Jurisdictional Immunity of States in International and Municipal Law* (Saarbrücken: Europa-Institut der Universität des Saarlandes, 1988) 14.

25 *La Forest J in United States of America v Public Service Alliance of Canada* [1992] 2 SCR 50, 89.

26 *ibid passim*.

27 See also n 3 above, 1580 *per* Lord Clyde.

28 As regards the terms 'acts *jure imperii*' and 'acts *jure gestionis*', Lord Clyde comments that the traditional Latin labels 'are convenient as words of reference but do not assist significantly in the application of the distinction'. *ibid* 1579.

29 Lord Clyde concedes that '[t]here is little if anything to be gained by trying to fit the case into a particular precedent or to devise categories of situations which may or may not fall on the one side of the line or the other.' *ibid* 1580.

A State is not immune as respects proceedings in respect of -

- (a) death or personal injury; or
- (b) damage to or loss of tangible property,

caused by an act or omission in the United Kingdom.

Apparently this section does not cover defamation claims. More clearly, in respect of Article 11 of the European Convention on State Immunity, which is plainly the basis of section 5, it has been explained: 'Where there has been no physical injury and no damage to tangible property, the article does not apply. This is the case, for example, as regards ... defamation.'³⁰

Admittedly, the possibility cannot be excluded that defamation cases are covered by other sections, not least because in ratifying the Convention the UK made the following declaration in accordance with Article 24(1)

[I]n cases not falling within Article 1 to 13 [ie Convention exceptions to immunity], their courts ... shall be entitled to entertain proceedings against another Contracting State to the extent that these courts are entitled to entertain proceedings against States not Party to the present Convention. This declaration is without prejudice to the immunity from jurisdiction which foreign States enjoy in respect of acts performed in the exercise of sovereign authority (*acta jure imperii*).³¹

On the face of it, however, this declaration does not mean that immunity is withheld from a foreign state for every act *jure gestionis*.³² Accordingly, it would not contradict this declaration to suggest that, even if the act at issue is an act *jure gestionis*, a foreign state is immune in defamation proceedings, which fall outside the scope of section 5. Indeed, as will be seen, Lord Millett's view comes close to this suggestion.

The plaintiff invoked, if statutory state immunity were to be applied, the so-called commercial transaction exception (section 3), the relevant part of which provides:

- (1) A State is not immune as respects proceedings relating to -
 - (a) a commercial transaction entered into by the State; ...
 - (b) [...]
- (2) [...]
- (3) In this section 'commercial transaction' means -
 - (a) any contract for the supply of goods or services;
 - (b) [...]
 - (c) any other transaction or activity (whether of a commercial, industrial, financial, professional or other similar character) into which a State enters or in which it engages otherwise than in the exercise of sovereign authority; ...

This section is couched in more extended terms than the corresponding provisions of the Convention. Thus, a commentator states that 'the breadth of the definition of "commercial transaction" in section 3 (...) makes it not inconceivable that immunity will in future be denied in [libel] cases like *Krajina v. Tass Agency*.'³³

³⁰ *Explanatory Reports on the European Convention on State Immunity and the Additional Protocol* (Strasbourg: Council of Europe, 1972) 20 (para 48).

³¹ Cmnd 7742 (1979) at 19.

³² See also *Kuwait Airways Corp v Iraqi Airways Co* [1995] 1 WLR 1147 (HL) 1171-1172 per Lord Mustill. But see *ibid* 1158 per Lord Goff.

³³ Sinclair, n 5 above, 261-262. In *Krajina v Tass Agency*, defamation proceedings against a state agency, immunity was granted pursuant to the principle of absolute immunity which the English courts then adopted. [1949] 2 All ER 274 (CA).

As regards this section, Lord Millett observed that 'the words "proceedings relating to" a transaction refer to claims arising out of the transaction, usually contractual claims, and not tortious claims arising independently of the transaction but in the course of its performance.'³⁴ From this standpoint, he denied the application of subsection 3(a) because *Holland v Lampen-Wolfe* does not relate to the contract between the university and the US, but to the memorandum. As regards subsection 3(c), he doubted that the writing and publication of the memorandum constituted an 'activity' within the meaning of this subsection, for '[t]he context strongly suggests a commercial relationship akin to but falling short of contract (...) rather than a unilateral tortious act.'³⁵

Persuasive though this opinion is, it is nothing but a dictum. The other Law Lords expressed no opinion about section 3. In any event, as Lord Millett admitted, subsection 3(c) would have been inapplicable in this case due to the finding that the memorandum was written and published 'in the exercise of sovereign authority'. Thus, no definite answer was provided to the question whether a unilateral tortious act, including defamation, can constitute an 'activity' and can be covered by section 3.

Further, it appears that the reasoning of the Law Lords could have a knock-on effect upon statutory state immunity in defamation proceedings. There is a ring of truth in the already quoted remark of Lord Millett to the effect that Parliament's intention to withhold immunity from members of visiting armed forces is unlikely in circumstances where immunity is granted to other state officials.³⁶ Another finding of the Law Lords is, as noted, that in proceedings relating to visiting armed forces no immunity is accorded for acts *jure gestionis*. These taken together, a presumption is that, while the State Immunity Act 'does not adopt the straightforward dichotomy between *acta jure imperii* and *acta jure gestionis*',³⁷ in so far as the latter is concerned Parliament made a provision for denying immunity without exception.

If defamation claims against a foreign state were to be covered by section 3 of the State Immunity Act, immunity would surely depend upon whether the act at issue is carried out 'in the exercise of sovereign authority' (subsection 3(c)). Then, it is manifest that the same problems arise as pointed out in respect of state immunity at common law: whether it is justified to treat victims of defamation differently in accordance with the distinction between sovereign acts and non-sovereign acts; how this distinction is applied in the context of defamation.

The writer is aware that a number of US courts have recently applied the commercial activity exception of the US Foreign Sovereign Immunities Act 1976 in defamation cases notwithstanding the exclusion of 'any claim arising out of ... libel, [or] slander' from the tort exception (s 1605(a)(5)(B)).³⁸ Yet, there is a difference between this US statute and the State Immunity Act. Under the latter, a separate entity as distinct from a state is immune in the first place only if 'the proceedings relate to anything done by it in the exercise of sovereign authority' (section 14(2)), whereas no similar qualification exists under the US statute (s 1603(b)). Consequently, under the US statute, if the commercial activity exception were not applied, a wide range of separate entities would enjoy immunity in defamation proceedings. This is not the case under the State Immunity Act.

34 n 3 above, 1587

35 *ibid.*

36 See the text accompanying n 6 above.

37 *Alcom Ltd v Republic of Colombia* [1984] 1 AC 580, 600 per Lord Diplock.

38 See eg *El-Hadad v Embassy of United Arab Emirates*, 69 F Supp 2d 69 (DDC 1999); *Alpha Therapeutic Corp v Nippon Hoso Kyokai*, 199 F3d 1078 (9th Cir 1999).

State immunity and the right to the court

As pointed out, the Law Lords held that the defendant in *Holland v Lampen-Wolfe* enjoyed immunity from suit. From the viewpoint of the right to the court, this conclusion is in marked contrast to the decision concerning the advocate's immunity, where Lord Millett thought that 'it is difficult to defend a blanket professional immunity in terms of the European Convention [on Human Rights].'³⁹ Since the immunity at issue was state immunity at common law, it is evident that this contrast is not due to the fact that, while the advocate's immunity was a purely common law principle, there is an Act of Parliament concerning state immunity. What then, brought about the difference between the advocate's immunity and state immunity?

Needless to say, if the right to the court is unrestricted, no immunity from suit can stand. Yet, at least in the framework of the European Convention on Human Rights, this right has not been considered so absolute as to reject any sort of immunity from suit.⁴⁰ In *Golder v United Kingdom*, while holding that Article 6(1) of the Convention 'embodies the "right to a court",' the European Court of Human Rights did not fail to add that 'the right of access to the courts is not absolute.'⁴¹ In subsequent cases, the Court has developed this qualification in the following terms:

[A] limitation [to the right of access to the courts] will not be compatible with Article 6 § 1 if it does not pursue a legitimate aim and if there is not a reasonable relationship of proportionality between the means employed and the aim sought to be achieved.⁴²

The Law Lords could have accepted or rejected the plaintiff's submission from this standpoint. Indeed, it was based upon this approach that the European Court of Human Rights in *Waite v Germany* considered the immunity of an international organisation compatible with Article 6(1).⁴³ Lord Millett, however, took a different approach:

At first sight [the right to a fair trial] may appear to be inconsistent with a doctrine of comprehensive and unqualified state immunity in those cases where it is applicable. But in fact there is no inconsistency. This is not because the right guaranteed by article 6 is not absolute but subject to limitations, nor is it because the doctrine of state immunity serves a legitimate aim.⁴⁴

He went on to say

It is because article 6 forbids a contracting state from denying individuals the benefit of its powers of adjudication; it does not extend the scope of those powers. ... It presupposes that the contracting states have the powers of adjudication necessary to resolve the issues in dispute. But it does not confer on contracting states adjudicative powers which they do not possess. State immunity, ... is a creature of customary international law ... It is not a self-imposed restriction on the jurisdiction of its courts which the United Kingdom has chosen to adopt. It is a limitation imposed from without upon the sovereignty of the United Kingdom itself.⁴⁵

Similarly, in concluding that 'the application of [state] immunity would [not] be unreasonable or disproportionate', Lord Clyde states that 'the present case is

³⁹ n 2 above, 624.

⁴⁰ See *X v Austria*, App No 3374/67 [1969] YECHR 246 (European Commission of Human Rights) concerning parliamentary immunity.

⁴¹ Ser A, no 18, 18 (paras 36 and 38).

⁴² *Ashingdane v United Kingdom*, Ser A, no 93, 24-25 (para 57). See also, as recent authority, *Osman v United Kingdom*, App No 23452/94 (2000) 29 EHRR 245, 315 (para 147).

⁴³ App No 26083/94, 6 BHRC 499 (1999), esp 510 (para 63) and 512 (para 73).

⁴⁴ n 3 above, 1588.

⁴⁵ *ibid.*

concerned with a ... situation where, as matter of the international relations between states, the domestic courts in the United Kingdom lack any jurisdiction to provide any remedy.⁴⁶

The opinion of the European Commission of Human Rights in *Vearncombe v United Kingdom and Germany*⁴⁷ is instructive in this respect. As regards the application against Germany, the facts of this case are fairly close to those of *Holland v Lampen-Wolfe*. The applicants brought an action to enjoin the British military authorities in Berlin from constructing and using a shooting range which would cause excessive noise nuisance. Under the Allied Kommandatura Law, the German courts are in principle prohibited from exercising criminal or civil jurisdiction over the Allied Forces, or in respect of acts arising out of or in the course of performance of duties or services with the Forces. The German court therefore held the action inadmissible due to lack of jurisdiction. The Commission observed that '[a]cts performed by organs of an occupying State (...) are generally attributable to this State and not to the occupied State', and decided that under these circumstances the application against Germany 'is incompatible *ratione personae* with the provisions of the Convention and must to this extent be rejected'.⁴⁸

The rules of international law, of which state immunity is one, have their principal function in the allocation of jurisdiction among states. Then, the approach that, in cases where state immunity is applicable, a forum state lacks jurisdiction to provide the right to the court seems analytically correct. If, in the context of state immunity, the right to the court is to be invoked, it is against a defendant state which enjoys immunity rather than against a forum state which grants immunity. For, 'where the jurisdictional protection of [a forum state's] courts is not available [due to the immunity of foreign states], it is always possible to seek a remedy before those States' own domestic courts'.⁴⁹

46 *Ibid* 1581.

47 App No 12816/87, 59 DR 186 (1989).

48 *Ibid* 193. See also *Spaans v Netherlands*, App No 12516/86, 58 DR 119 (1988); concurrent opinion of Mr K. Herndl in *Beer and Regan v Germany*, App No 28934/95 [1998] EHRLR 478 (European Commission of Human Rights) 479.

49 *Nacci v Bari Institute*, 114 ILR 539 (Italy, Court of Cassation, Plenary Session, 1994) 554.

Sale and Lease-Back Agreements in a World of
Title Relativity:
*Michael Gerson (Leasing) Ltd v Wilkinson and
State Securities Ltd*

*Janet Ulph**

Introduction

Where there is misconduct, the courts are often faced with a choice between conflicting proprietary claims made by those who have innocently become caught up in the rogue's activities. Very rarely, there is a windfall profit and the court may have the relatively pleasant task of choosing who has the better claim to it;¹ however, usually the rogue is insolvent and one party may suffer as a result.²

In an insolvency, the first hurdle for a claimant will be to establish proprietary rights in relation to identifiable property in order to elevate himself above the position of an unsecured creditor. Yet there is always a danger that a claimant can be *divested* of these proprietary rights if another party can also establish proprietary rights over the same property and can successfully rely upon one of the exceptions to the *nemo dat* principle, that no-one can transfer a better title than they have themselves. The recent Court of Appeal decision in *Michael Gerson (Leasing) Ltd v Wilkinson and State Securities Ltd*.³ is a significant case, which extends the protection offered to a second purchaser by the 'seller-in-possession' exception and exposes finance companies to unanticipated risks in relation to goods which they own but which are currently subject to sale and leaseback agreements.

Background

The essential facts were that the seller, Emsshelf IX Ltd. ('Emsshelf'), sold certain plant and machinery to Michael Gerson (Leasing) Ltd. ('Gerson'), a finance company, under a sale and leaseback agreement under which Emsshelf remained in physical possession of this equipment. A year later, Emsshelf sold the very same equipment to State Securities Ltd. ('State'), which was another finance company, under a second sale and leaseback agreement. When Emsshelf became insolvent, Gerson sued State in conversion and State successfully pleaded in defence that the 'seller-in-possession' exception, contained in section 8 of the Factors Act 1889 and section 24 of the Sale of Goods Act 1979, applied.

Section 24 of the Sale of Goods Act provides:

Where a person having sold goods continues or is in possession of the goods, or of the documents of title to the goods, the delivery or transfer by that person ... of the goods or documents under any sale, pledge, or other disposition thereof, to any person receiving the same in good faith and without notice of the previous sale, has the same effect as if the

* Department of Law, University of Durham.

1 *Foskett v McKeown* [2001] 1 AC 102, 106 per Lord Browne-Wilkinson.

2 *Central Newbury Car Auctions Ltd v Unify Finance Ltd* [1957] 1 QB 371, 379 per Denning LJ.

3 [2000] 3 WLR 1645.

person making the delivery or transfer were expressly authorised by the owner of the goods to make the same.

Thus the seller-in-possession exception may be called into play wherever goods have been purchased but the seller remains in possession of them and subsequently delivers them to a second purchaser or pledgee who can show that the goods were received in good faith. On the facts, the seller, Emsshelf had remained in continuing physical possession of the goods. The difficult issue was whether the second purchaser, State, could be said to have received them. State never physically received the goods and it therefore had to argue that, according to the provisions of the sale and leaseback agreement, there had been a notional 'constructive' delivery of the goods by Emsshelf before the goods were leased back.

The debate regarding constructive delivery

The question of whether physical delivery of the goods into the hands of a second purchaser is necessary in order to satisfy the statutory provision, or whether proof of constructive delivery is enough, has been a controversial one. Interestingly, in *Michael Gerson*, there was no battle over this matter in the Court of Appeal: counsel for the claimant conceded that, for the purposes of the appeal, constructive delivery sufficed.⁴ However, as the concession was so carefully limited, the question could be raised afresh in the future. The matter therefore deserves analysis.

The authorities on this issue have been in disarray for some time. The main difficulty lay with the first instance decision of *Nicholson v Harper*,⁵ in which North J stated that there must be actual physical delivery of the goods after the second sale or pledge in order to satisfy the statutory wording of the seller in possession exception. However, there was also the case of *NZ Securities & Finance Ltd v Wrightcars Ltd*,⁶ in which an argument that constructive delivery of goods had been made under a sale and leaseback agreement was rejected and it was held that there must be a physical delivery of the goods to the purchaser. In coming to that decision, the New Zealand court relied upon *Nicholson v Harper* and the Australian case of *Bank of New South Wales v Palmer*.⁷ In the latter case, Helsham J suggested that the view which he had formed that the seller-in-possession exception required physical delivery was supported by the statutory word 'receiving.'

Yet the decision in *Nicholson v Harper* could be described as deeply flawed. There was no possibility on the facts of the case of actual physical delivery because the second pledgee was already in possession of the goods before the pledge in his capacity as warehouseman. This fact was brought to North J.'s attention but the argument that there had been a constructive delivery to the pledgee was rejected without discussion and without any reference to past case law. There were a number of cases decided before *Nicholson v Harper* which North J should have considered. In *Elmore v Stone*,⁸ for example, the claimant had sold two horses to

4 [2000] 3 WLR 1645, 1649 E-H, per Clarke LJ. At first instance, Michael Kershaw QC, sitting as a judge of the High Court, had held that delivery can be constructive for the purposes of the Sale of Goods Act 1979, s. 24: *Michael Gerson (Leasing) Ltd v Michael Wilkinson and State Securities Ltd* (1998) Unrep. 6 October, QBD (Liverpool).

5 [1895] 2 Ch. 415.

6 [1976] 1 NZLR 77 (Supreme Court in Wellington).

7 [1970] 2 NSWLR 532 (Supreme Court of New South Wales); overruled by *Gamer's Motor Centre (Newcastle) Proprietary Ltd v Natwest Wholesale Australia Proprietary Ltd* (1987) 163 CLR 236.

8 (1809) 1 Taunt 458.

the defendant but had not been paid. Although there was no written memorandum of the sale as required by the Statute of Frauds, the agreement would still be enforceable by the claimant if delivery to the buyer could be established. Mansfield CJ held that, where the seller had agreed to look after the horses for the benefit of the buyer by keeping them at livery after the sale, there had been a constructive delivery of the horses to the buyer. He recognised that constructive delivery can take various forms and gave the example of where the key of a warehouse in which the goods are stored is delivered to the buyer.⁹ One might add, as another example, that a bill of lading is the symbol of property in goods and its delivery is treated as delivery of the goods themselves.¹⁰ There are other cases in which physical delivery has not been insisted upon.¹¹ However, in particular, there have been two persuasive modern authorities concerning another exception to the *nemo dat* rule: in both the first instance English decision of *Forsythe International (UK) Ltd v Silver Shipping Co Ltd and Petroglobe International Ltd (The Saetta)*¹² and the decision of the High Court of Australia in *Gamer's Motor Centre (Newcastle) Proprietary Ltd v Natwest Wholesale Australia Proprietary Ltd*,¹³ it was accepted that proof of constructive delivery would be enough to satisfy the buyer in possession exception.

The concession made in *Michael Gerson* that constructive delivery sufficed, which was welcomed by Clarke LJ,¹⁴ was surely correct in principle: arguably *Nicholson v Harper* and subsequent authorities¹⁵ would not have stood up to extensive scrutiny because the courts in those cases had not engaged in a detailed review of past case law. Furthermore, the emphasis placed upon the word 'receiving' by Helsham J in *Bank of New South Wales v Palmer*¹⁶ could be seen as self-serving. For example, the word 'taking' can be found in the mercantile agency exception and no significance is attached to it. It is submitted that, if the defendant can prove constructive delivery, then he can properly be described as having received the goods.

Defining constructive delivery: transfer of control

It is delivery to the second purchaser or pledgee, rather than the antecedent contract, which has the effect of divesting the first purchaser of his proprietary rights under the seller-in-possession exception. But what exactly is needed for

9 *Ibid* 460. See further *Ward v Turner* (1751) 2 Ves Sen 431, 443 *per* Lord Hardwicke; *Dublin City Distillery Ltd v Doherty* [1914] AC 823, 843 *et seq.*, *per* Lord Atkinson, 852, *per* Lord Parker; *Wrightson v McArthur and Hutchisons (1919) Ltd* [1921] 2 KB 807, 816–817, *per* Rowlatt J; *Akron Tyre Co Pty Ltd v Kittson* (1951) 82 CLR 477, 494 *per* Williams and Kitto JJ; *Official Assignee of Madras v Mercantile Bank of India Ltd* [1935] AC 53, 58–59 *per* Lord Wright.

10 *Meyerstein v Barber* (1866) LR 2 CP 38, 48 *per* Erie CJ, 50 *per* Willes J; *Official Assignee of Madras v Mercantile Bank of India Ltd* [1935] AC 53, 59 *per* Lord Wright.

11 *Marvin v Wallis* (1856) 6 EL & BL 726, 25 LJQB 369; *Whitehouse v Frost* (1810) 12 East 614; *Ministry for Supply & Development v Servicemen's Co-operative Joinery Manufacturers Ltd* (1951) 82 CLR 621; *Re Morrison* (1905) 25 NZLR 532; *Re Stoneham* [1918–1919] All ER 1051; *Thomas v The Times Book Co Ltd* [1966] 2 All ER 241.

12 [1994] 1 WLR 1334.

13 (1987) 163 CLR 236.

14 [2000] 3 WLR 1645, 1649 E–H, *per* Clarke LJ.

15 *Bank of New South Wales v Palmer* [1970] 2 NSWLR 532; *NZ Securities & Finance Ltd v Wrightcars Ltd* [1976] 1 NZLR 77.

16 [1970] 2 NSWLR 532, 535 (Supreme Court of New South Wales); overruled by *Gamer's Motor Centre (Newcastle) Proprietary Ltd v Natwest Wholesale Australia Proprietary Ltd* (1987) 163 CLR 236.

constructive delivery? There is statutory guidance in relation to the notion of delivery itself: according to section 61 (1) of the Sale of Goods Act 1979, unless the context suggests otherwise, 'delivery' is the '*voluntary transfer of possession from one person to another*.' The transfer must be voluntary in the sense that a deliberate step must be taken.¹⁷ When one then turns to consider possession, statutory guidance on constructive possession is provided by section 1(2) of the Factors Act which states that, 'A person shall be deemed to be in possession of goods or of the documents of title to goods, where the goods or documents of title are in his actual custody or are held by any other person subject to his control or for him on his behalf.' It is clear from this section that it is wrong to focus solely upon transfer of physical possession; the key is the transfer of control.

Yet commercial arrangements can be complex and gauging the nature and extent of control exercised by one party may not be straightforward. One of the difficulties facing the Court of Appeal, in determining if and when control could be said to have been transferred under a sale and leaseback arrangement, was that for many years the question has been obscured by the shadow cast by *Nicholson v Harper*. However, the court confirmed that there may be constructive delivery where there is evidence of a change in the character of possession.¹⁸ This would occur where there has been an attornment: for example, where the seller, although remaining in possession, acknowledges that he holds the goods on account of the buyer;¹⁹ or where a warehouseman does so,²⁰ or indeed where the buyer already had possession but now acknowledges that he holds them for his own benefit.²¹

In coming to its decision, the Court of Appeal considered *Gamer's Motor Centre (Newcastle) Proprietary Ltd v Natwest Wholesale Australia Proprietary Ltd*.²² The erudite leading judgment given by Mason CJ in *Gamer* was largely taken up with justifying the principle that constructive delivery was sufficient. If not for the irritant of *Nicholson v Harper*, the High Court of Australia in *Gamer* could have simply concentrated on the crucial question of whether constructive delivery could be found to have taken place on the facts. An agreement had been made by Gamer to sell cars to a dealer subject to a reservation of title clause. The dealer took possession of the cars and resold them to a finance company, Natwest. Natwest never took actual physical possession of the cars; instead, the dealer retained them. Nevertheless, Natwest was able to rely upon the buyer in possession exception. The High Court of Australia held that there had been constructive delivery when the dealer sent a delivery receipt to Natwest giving particulars of the transaction and confirming in effect that Natwest was the owner.²³ A criticism which had been

17 *Forsythe International (UK) Ltd v Silver Shipping Co Ltd and Petroglobe International Ltd (The Saetta)* [1994] 1 WLR 1334, affirmed by Clarke LJ in *Michael Gerson (Leasing) Ltd v Wilkinson and State Securities Ltd* [2000] 3 WLR 1645.

18 [2000] 3 WLR 1645, 1655 D-E, per Clarke LJ; 1668H-1669A per Pill LJ. See F. Pollock and R.S. Wright, *An Essay on Possession in the Common Law* (Oxford: Clarendon Press, 1888) 72-73; *Bowstead & Reynolds on Agency* (Sweet & Maxwell, 16th edn., 1996, para 8-170).

19 *Marvin v Wallis* (1856) 6 El. & Bl. 726; *Elmore v Stone* (1809) 1 Taunt 458, *Hurry v Mangles* (1808) 1 Camp 452; *Whitehouse v Frost* (1810) 12 East 614; *Castle v Swoode* (1861) 6 H & N 828; contrast, on different facts, *Dublin City Distillery Ltd v Doherty* [1914] AC 823; *Townley v Crump* (1835) 4 Ad & B 58; *Carter v Toussaint* (1822) 5 B & Ald 855.

20 *McEwan v Smith* (1849) 2 H L C 309, 325 per The Lord Chancellor; *Dublin City Distillery Ltd v Doherty* [1914] AC 823, 847 per Lord Atkinson, 852, per Lord Parker of Waddington; *Official Assignee of Madras v Mercantile Bank of India Ltd* [1935] AC 53, 58-59 per Lord Wright. See further, *Laurie and Morewood v Dudin & Sons* [1926] 1 KB 223; *Maynegram Pty Ltd v Compafina Bank* [1982] 2 NSWLR 141.

21 *Re Morrison* (1905) 25 NZLR 532.

22 (1987) 163 CLR 236.

23 *ibid* 250 per Mason CJ 263 per Dawson J.

made of the decision in *Gamer* was that a simple delivery note would not ordinarily amount to acknowledgment of change of control.²⁴ Clarke LJ expressly dealt with this issue in *Michael Gerson*. He drew upon the report of the facts of *Gamer* at a lower appellate level²⁵ and concluded that the delivery order in *Gamer* had in fact been more than a simple receipt. The delivery order expressly provided that Natwest could recover the goods without notice and this statement evidenced the transfer of control to Natwest.

The Court of Appeal in *Michael Gerson* then had to consider the particular obligations imposed by a sale and leaseback agreement. There would have had less difficulty if the facts had been otherwise and there had been an ordinary sale followed by a later request from the seller to borrow the goods for a specified period of time. These were the facts in *Marvin v Wallis*,²⁶ and it was held that, as the seller had acknowledged the buyer's title and merely held the goods as a bailee, there had been constructive delivery of the goods to the buyer sufficient to satisfy the Statute of Frauds. *Michael Gerson* was different: there was one entire transaction. It was argued on behalf of the claimant that Emsheff never lost control of the goods under such an agreement: there was not a moment in time when State could truly be said to be the owner with a right to insist on actual physical delivery of the goods and with the ability to choose whether to lease the goods to Emsheff or not.²⁷ This argument had considerable force because of the fact that, once the lease took effect, it would be Emsheff who would be treated in law as having possession of the goods.²⁸ However, the court rejected the argument. In effect, the sale and leaseback agreement was taken as a genuine description of what the parties intended. The court considered the terms of the lease and reasoned that State would not have been entitled to declare that it was the owner of the goods or to fulfil its obligations under the lease (such as delivery to Emsheff) if it did not have constructive possession of the goods in law at some point. It was concluded that there had been a constructive delivery to State followed by a redelivery to Emsheff. Clarke LJ admitted that he was influenced by commercial considerations in coming to his decision: he thought that sale and leaseback agreements should be brought within the scope of the seller in possession exception.²⁹

Implications and critique of the decision

Judged purely from a sale of goods perspective, the decision in *Michael Gerson* can be welcomed. It provides useful guidance on what amounts to constructive delivery by focusing upon control: sale and leaseback agreements were analysed as involving a transfer of control over the goods to the purchaser followed by a retransfer of control to the lessee. Pill LJ suggested that there was a danger of creating fine distinctions between sale and leaseback agreements and other transactions otherwise.³⁰ If one accepts that sale and leaseback agreements are

24 See K.C.T. Sutton, *Sales and Consumer Law* (LBC Information Services, 1995) para 17.21.

25 [1985] 3 NSWLR 475 (Court of Appeal of New South Wales); see [2000] 3 WLR 1645, 1650 G-H per Clarke LJ.

26 (1856) 6 El. & Bl. 726, 25 LJQB 369 (cited in the Court of Appeal as *Marvin v Wallace*).

27 [2000] 3 WLR 1645, 1655 F-G, 1669 B-C.

28 See *Anglo-Irish Asset Finance v DSG Financial Services* [1995] CLY 4491, where it was held that, in the context of the seller in possession exception, it was the hirer, rather than the owner, who had constructive possession of cars let on hire-purchase.

29 [2000] 3 WLR 1645, 1657 A-B. Pill LJ preferred to express no view on this point: *ibid* 1669 D.

30 [2000] 3 WLR 1645, 1669D.

more than artificial credit arrangements and genuinely do involve a sale of goods followed by a lease, then they are difficult to distinguish from situations where the seller, a moment after the sale, asks to borrow the goods. Indeed, in *Marvin v Wallis* itself, the request by the seller to borrow the horse followed within seconds of completion of the sale.

Although it was conceded that constructive delivery was sufficient for the 'seller-in-possession' exception, the decision in *Michael Gerson* can be used to add further weight to the argument that, in principle, constructive delivery should suffice for both the 'seller-in-possession' and 'buyer-in-possession' exceptions. Actual physical delivery has never been required for any of the other exceptions to the *nemo dat* rule and, although each exception has its own particular conditions which have to be met, harmony on basic principles is desirable. Moreover, the Court of Appeal's decision in *Michael Gerson* can be seen as in tune with the historical development of the seller-in-possession exception: Parliament was concerned to protect purchasers where, to all outward appearances, the seller continued to appear as owner of the goods because he still remained in possession of them.³¹

The Court of Appeal's analysis of the law creates greater harmony within the Sale of Goods Act 1979 itself. The concept of constructive delivery and the question of relinquishing control is highly significant in relation to transfer of property. For example, evidence that the seller intended to retain control over the goods after the contract is made will suggest that property in specific goods was not intended to pass at the time of the contract.³² There is also a linked relationship between control and delivery on the one hand and the transfer of property in the goods by appropriation with the other party's assent on the other: this is true not only in relation to the purchase of a specific quantity of goods from a larger bulk³³ but also in relation to the purchase of generic goods which have been ascertained.³⁴ As *Re Stapylton Fletcher Ltd*³⁵ illustrates, there may be a constructive delivery and the property in goods may be transferred to a buyer even where, because the contract is one of sale and storage, the goods are never physically handed over and returned but simply moved to another warehouse to be stored until required.

Yet some difficulties remain. One argument made on behalf of the claimant was that the defendant did not have possession within the meaning of section 1(2) of the Factors Act 1889. This was rejected on the basis that constructive possession would suffice.³⁶ It is submitted that this aspect of the decision is uncontroversial.³⁷

31 *Pacific Motor Auctions Pty Ltd v Motor Credits (Hire Finance) Ltd* [1965] AC 867, 886 per Lord Pearce on behalf of the Privy Council.

32 *Ministry for Supply & Development v Servicemen's Co-operative Joinery Manufacturers Ltd* (1951) 82 CLR 621, 634-635 per Latham CJ, 641 per Williams J.

33 See Sale of Goods Act 1979, s 61 (1), as amended. 'delivery' means voluntary transfer of possession from one person to another, except that in relation to sections 20A and B above it includes such appropriation of goods to the contract as results in property in the goods being transferred to the buyer. See further, L. Gullifer, *Constructive Possession after the Sale of Goods (Amendment) Act 1995* [1999] LMCLQ 93, 104.

34 Sale of Goods Act 1979, s 18 r 5 (1) (2); *Carlos Federspiel & Co SA v Charles Twigg & Co Ltd* [1957] 1 Lloyd's Rep 240, 255 per Pearson J.

35 [1994] 1 WLR 1181, 1200 per Judge Paul Baker QC.

36 [2000] 3 WLR 1645, 1656 F-H.

37 On constructive possession, see *City Fur Manufacturing Co Ltd v Furenbond (Brokers) London Ltd* [1937] 1 All ER 799; *Forrythe International (UK) Ltd v Silver Shipping Co Ltd and Petroglobe International Ltd* (The Saetta) [1993] 2 Lloyd's Rep 268; *Four Point Garage Ltd v Carter* [1985] 3 All ER 12; *Beverley Acceptances Ltd v Oakley* [1982] RTR 417; *Lowther v Harris* [1927] 1 KB 393. A person will have constructive possession of goods as holder of a document of title: *Cahn and Mayer v Pockett's Bristol Channel Steam Packet Co Ltd* [1899] 1 QB 643; *Lickbarrow v Mason* (1794) 5 TR 683; *Kum v Wah Tat Bank Ltd* [1971] 1 Lloyd's Rep 439. See further, J. Ulph, 'Conflicts of Title and

However, somewhat surprisingly, Clarke LJ appeared reluctant to accept the wider argument that the Sale of Goods Act and the Factors Act should be read together as a code and appeared ready to accept that there might be differences of interpretation between them. This raises uncertainty where none existed before. A good-faith purchaser can rely on either set of provisions which, apart from one minor detail, are the same.³⁸ The two statutes are frequently read together:³⁹ in particular, even where the Sale of Goods Act is pleaded, a courts must use the same generous definition of a document of title which is to be found in the Factors Act.⁴⁰ A lack of consistency between the two statutes would be completely undesirable in an area of law which is already highly technical.

There is one outstanding issue concerning the seller-in-possession exception. It was established by the Privy Council in *Pacific Motor Auctions Pty Ltd v Motor Credits (Hire Finance) Ltd*⁴¹ that a seller has the power to pass a good title to the second buyer provided he remains in continuous physical possession of the goods regardless of the character of the possession. Will constructive possession by the seller suffice in these circumstances?⁴² It is submitted that it should where, for example, the seller does not have actual physical possession of the goods because they are stored in a warehouse.⁴³

Yet it has to be said that the decision in *Michael Gerson* may not receive a warm welcome from insolvency practitioners. One weakness was that, as there was one transaction, it was difficult to pinpoint exactly when the goods were delivered to State. The court acknowledged the problem but Clarke LJ sidelined this difficulty by suggesting that it was unnecessary to identify a moment at which the goods were delivered.⁴⁴ This is unsatisfactory. If a decision in relation to proprietary rights hangs upon whether constructive delivery has taken place or not, then the court should determine the exact time when such delivery occurred. Traders need to know when proprietary rights have been vested or divested. Timing is everything where there is an insolvency.⁴⁵ Arguably the Court of Appeal in *Michael Gerson* should have gone further and expressly stated that constructive delivery of the goods to State took place when the agreement was made: for a moment in time, State could be described as the owner before the lease took effect and redelivery of the goods occurred.

An interesting aspect of the decision relates to the use of sale and leaseback arrangements as a form of secured lending. These agreements have always been vulnerable to attack by creditors in a company's insolvency: if it can be shown that

the Obligations of the Seller' in E. McKendrick (ed), *Sale of Goods* (London: LLP, 2000) paras 5-040, 5-060 and 5-083.

38 Sale of Goods Act 1979, s 21(2) provides that the provisions of the Factors Act are not excluded.

39 *Cahn and Mayer v Pockett's Bristol Channel Steam Packet Co. Ltd* [1899] 1 QB 643, 652, per A.L. Smith LJ; *City Fur Manufacturing Co Ltd v Fureenbond (Brokers) London Ltd* [1937] 1 All ER 799, 802 per Branson J; *Forsythe International (UK) Ltd v Silver Shipping Co Ltd and Petroglobe International Ltd (The Sactta)* [1993] 2 Lloyd's Rep 268, 275 per Clarke J.

40 Sale of Goods Act 1979, s. 61(1). See further Halsbury's *Laws of England* vol 41, para 751 n 8.

41 [1965] AC 867.

42 J. Mo, "Physical Possession" versus "Constructive Delivery" in Section 28 of the Sale of Goods Act 1923 (NSW) - "Harmonisation" of *Pacific Motor Auctions and Gamer's Case* (1993) 16(2) UNSW LJ 513, 515, 525

43 This issue was raised by the High Court in *Gamer's Motor Centre (Newcastle) Pty Ltd v Natwest Wholesale Australia Pty Ltd* (1987) 163 CLR 236; the judges in the majority were content to give 'delivery' a different meaning from possession if this was necessary: *ibid* Mason CJ at 248-249; 254 per Brennan J, 260 per Dawson J.

44 [2000] 3 WLR 1645, 1655 H.

45 See, for example, *Neste Oy v Lloyd's Bank plc* [1983] 2 Lloyd's Rep 658.

an agreement is in substance a disguised mortgage to secure a loan then it is unenforceable unless it has been registered.⁴⁶ It is a question of looking at the parties' intentions and the contractual terms to decide whether the sale was genuine or not.⁴⁷ This line of attack would have been unattractive to the claimant given that its own right to possession rested on the validity of its sale and leaseback agreement. In any event, the courts have tended to take a generous approach in upholding these agreements. However, the decision in *Michael Gerson* provides a cautionary reminder to finance companies that the flexibility which makes these agreements so appealing may be bought at the expense of safety regarding their security. The defendant succeeded in divesting the claimant of its proprietary rights by using an exception to the *nemo dat* rule. Yet the effect of the decision is that the defendant's proprietary rights could just as easily have been divested in turn if a third sale and leaseback agreement had been made. Owners operate in a world of title relativity;⁴⁸ there is always a risk that another person may come forward with a better right to possession.

A sale and leaseback agreement is a useful method by which an ailing trader can obtain a quick injection of capital to keep a business going. Will the decision in *Michael Gerson* affect its popularity? Ordinarily, it would not be feasible for finance companies to take physical possession of goods before leasing them back in order to avoid the situation where the lessee remains as a 'seller in possession.' It is more likely that finance companies will continue to make use of sale and leaseback agreements but will factor in this additional risk by offering to purchase at a lower valuation. Ironically, therefore, the court's decision to take a generous and pragmatic attitude to sale and leaseback arrangements, enabling innocent finance companies to take advantage of the 'seller-in-possession' exception, may tend to discourage their use. Yet the real problem lies in a more general malaise in the law of credit and security: years ago, in *Moorgate Mercantile Co. Ltd v Twitchings*,⁴⁹ Lord Wilberforce bewailed the lack of a proper registration system of movables⁵⁰ to protect innocent purchasers and pointed out that the absence of such a system provided a fertile ground for fraud. Law reform in this area is now overdue.

46 Companies Act 1985, s 395. Written agreements made by individuals may require registration as a bill of sale under the Bills of Sale Acts 1878–1891. See R.M. Goode, *Commercial Law* (Penguin Books, 2nd edn, 1995) 652–653.

47 *Re Curtain Dream plc* [1990] BCLC 925, 935 per Knox J; *Welsh Development Agency v Export Finance Co Ltd* [1991] BCLC 148, 154 per Dillon LJ, 186–187 per Staughton LJ. This was one of the main issues at first instance: the defence was that the claimant did not have a proprietary interest because its sale and leaseback agreement was merely a loan secured by a charge. This argument was rejected: *Michael Gerson (Leasing) Ltd v Michael Wilkinson and State Securities Ltd* (1998) Unrep. 6 October, QBD (Liverpool).

48 M.G. Bridge, 'The Title Obligations of the Seller of Goods' in N. Palmer and E. McKendrick (eds), *Interests in Goods* (London: LLP, 2nd edn, 1998) 303–304.

49 [1977] AC 890, 901.

50 See, in particular, the Crowther Committee's *Report of the Committee of Consumer Credit*, Cm 4596 (1971).

Only a Matter of Time

Nicole Busby*

Introduction

The European Court of Justice has ruled that the limitation of retrospective membership of an occupational pension scheme is contrary to the principle of effectiveness. The two-year limitation currently provided by UK law severely reduces the practical effect of the right to claim retroactive membership under European Community law. The Court found that the six-month time limit for the commencement of legal proceedings was not prohibited by application of the principle of effectiveness as, on the contrary, such time limits assist in the provision of legal certainty. The development of European law in this respect will be explored and the practical impact of the Court's ruling considered.

Background to the case and preliminary ruling

In a consistent application of Community law, the European Court of Justice (ECJ) has confirmed that membership of an occupational pension scheme does fall within the scope of Article 119 (now Article 141) of the EC Treaty, thus conveying a right of membership to part-time workers. This has been the position since the Court's earlier judgments in Case C-57/93 *Vroege v NCIV Instituut voor Volkshuisvesting – Stichting Pensioenfonds*¹ and Case C-128/93 *Fisscher v Voorhuis Hengelo – Stichting Bedrijfspensioenfonds voor de Detailhandel*.² In considering these cases, the Court found that the exclusion of part-time workers from such schemes amounted to indirect discrimination if it affected a greater number of women than men and could not be objectively justified.

In requesting a preliminary ruling under Article 177 (now Article 234) in the present case, the House of Lords sought to establish the compatibility of certain time limits set down in domestic legislation with actions brought under Article 119. The main proceedings concerned a range of claims (22 in total) brought by private and public sector employees and ex-employees following the *Vroege* and *Fisscher* rulings. In the earlier cases, the Court had held that the limitation of effects in time set down in the *Barber* judgment,³ which is now enshrined in the Protocol attached to the Treaty on European Union, did not apply to claims concerning exclusion from occupational pension schemes on the grounds of part-time status. Membership in such cases could be back-dated to 8 April 1976, this being the date on which the direct effect of Article 119 was established in *Defrenne v SABENA II*⁴ (referred to below as '*Defrenne II*'). Following these rulings 60,000 UK employees and former employees, who had previously been excluded from

* University of Stirling.

1 [1994] ECR I-4541.

2 [1994] ECR I-4583.

3 Case C-262/88 [1990] ECR I-1889.

4 Case 43/75 [1976] ECR 455.

contracted-out pension schemes on the grounds of their part-time status, initiated proceedings. The House of Lords sought clarification of the conciliation of these procedural rules with Community law to enable national courts to decide on the admissibility of the claims before them and to determine the periods of time for which claimants could seek retroactive membership of occupational pension schemes.

The procedural rules in question were set down in the Equal Pay Act 1970 (EqPA) and the Occupational Pension Schemes (Equal Access to Membership) Regulations 1976 ('the Occupational Pension Regulations'). The first provision, contained in section 2(4) of the EqPA, states that claims which seek to establish the operation of an equality clause will be time-barred if they are brought more than six months after the cessation of employment. This is referred to below as the 'six-month rule'. The second provision, which amended section 2(5) of the EqPA, applies a two-year limitation on the retroactive application of membership of an occupational pension scheme. Under the Regulations, the period for which retrospective admission to such a scheme applies is backdated two years prior to the claim being made.

In light of the limitations that these procedural rules confer on the application of Community law, it was contended that they had the effect of rendering the right to retroactive membership of an occupational pension scheme impossible or excessively difficult. The Birmingham Employment Tribunal in its decision of 4 December 1995, found that these procedures were compatible with European Community law. This decision was subsequently upheld by the Employment Appeal Tribunal and the Court of Appeal. However, the House of Lords, being the court of last instance, felt itself bound to refer the case to the European Court for consideration of the procedures in the context of the relevant European provisions.

The judgment of the European Court of Justice

In giving consideration to the compatibility of the 'six-month rule' with European provisions, the Court referred to the 'principle of effectiveness'. This seeks to ensure that procedural rules laid down by member states do not have the effect of rendering rights conferred by Community law ineffective in practice. With reference to the setting of time limits, the Court stated that where such time limits were reasonable, it was settled in case law that they represented an application of the principle of legal certainty by preventing administrative decisions from being vulnerable to legal challenge indefinitely.⁵ On this basis, and following the reasoning of Advocate-General Léger,⁶ the Court held that the six-month limitation provided by section 2(4) of the EqPA '... does not render impossible or excessively difficult the exercise of rights conferred by the Community legal order and is not therefore liable to strike at the very essence of those rights.'⁷

The Court then considered the limitation imposed by the application of the Occupational Pension Regulations that the pensionable service of a claimant is restricted to the two years prior to the date of her claim for the purposes of calculating her entitlement. The question before the Court related to whether

5 See Case 33/76 *Rewe* [1976] ECR 1989; Case C-261/95 *Palmarini v Istituto Nazionale della Previdenza Sociale* [1997] ECR I-4025.

6 In his opinion delivered on 14 September 1999. See also A-G Léger's opinion and the judgment of the ECJ in the related case C-326/96 *Levez v Jennings (Harlow Pools) Ltd* (1998) ECR I-7835.

7 *id* at 34.

Community law precluded such a rule on the basis that its application would render the conference of retroactive membership of an occupational pension scheme impossible in practice. The Court was guided by its earlier ruling in Case C-246/96 *Magorrian and Cunningham v Eastern Health and Social Services Board*,⁸ in which it held that the principle of effectiveness precluded the application of the equivalent Northern Irish regulation. It was acknowledged that, although application of the two-year time limit did not completely deprive the claimants of their rights, such restriction of access to membership did nonetheless have the effect of severely reducing the practical effect of the right to claim retroactive membership of an occupational pension scheme.

Practical effects of the ruling

The effect of this ruling on UK law will be the removal of the two-year limitation for equal pay claims concerning retroactive membership of occupational pension schemes. In future, it is likely that such claims will be made back to 1976 – the date of the judgment in *Defrenne II*. This is in line with the approach set out in the Court's earlier rulings in the *Vroege*⁹ and *Fisscher*¹⁰ cases. Given the ECJ's apparently clear ruling on this particular aspect, it appears that the Court's treatment of the Preston claim is likely to have a very substantial effect. However, there are certain other factors at play which look likely to severely diminish the usefulness of this decision for those both directly involved in the test cases and others awaiting guidance on how best to advance related claims.

The main area of difficulty arises from the distinction between the inter-related legal concepts of *time limits* and *limitations in time*. The former are legislative provisions which set out the time in which something must be done, represented in this case by the six-month time limit provided for by the EqPA. The latter relate to the limit imposed on the periods of service that may count towards the accrual of a benefit in the approximation of a remedy, such as the two-year limitation contained in the Regulations. Although theoretically diverse, the concurrent application of such provisions has the practical effect of fudging such apparent distinction. This is because the operation of a time limit imposed on the right to make a claim will depend on the duration of the limitation in time which will, in turn, have a direct effect on the availability of a remedy. In *Preston*, the fact that all of the appropriate pension schemes were amended during the years 1985 and 1996 in order to admit part-time workers as members will, ironically, present insurmountable difficulties for many women. Those whose employment has ceased on the grounds of retirement (or for some other reason) following their admittance to the appropriate scheme, will have been prevented from initiating proceedings on the grounds of the application of the two-year limitation. With respect to such claims, this ruling changes nothing as failure to initiate proceedings in the six months following cessation of the employment will render any potential claim time-barred despite the otherwise potentially retroactive nature of the claim.

It is true that some of the claimants represented in the present case have initiated proceedings timeously, acting on advice given by their trade unions. However, it is inestimable how many others have been denied the opportunity to claim retroactive

⁸ [1997] ECR I-4025.

⁹ n 1 above, para 32.

¹⁰ n 2 above, para 28.

membership of an occupational pension scheme from which they were refused membership on discriminatory grounds and now find their claims to be time-barred. This appears particularly unjust as their failure to instigate proceedings arose due to the apparent lack of a suitable legal basis at the time of their cessation of employment.

The maintenance of the six-month rule, beyond which such claims will be time-barred, will, if sustained, severely limit the otherwise wide application of this judgment. The ECJ dealt with the two rules separately in line with the terms in which the referred questions were framed. However, for the reasons set out above, it is necessary to apply an integrated approach in order to analyse the likely practical impact of the ruling. Of course the House of Lords may still choose to disapply the six month rule to claims of this nature, but it is expected that its judgment will not be given until Spring 2001 adding to the legal uncertainty that the European Court had sought to avoid.

In light of this inconclusiveness, it is worth paying due consideration to the effect of the ECJ's ruling on the different categories of claims that would potentially benefit from the removal of the two-year limitation. This will be achieved by using a crude distinction consisting, in the first instance, of those claims that will be time-barred if the six-month time limit is sustained and, secondly, those where claimants have instigated proceedings within the current time constraint.

Application of the six-month time limit

Whatever the outcome of the House of Lords' judgment with respect to the sustainability of the six-month time limit, the European Court's treatment of this aspect of the present case merits further consideration. The general rule is that, where appropriate, domestic rules of limitation will apply to cases based on European Community law,¹¹ unless such application will make it impossible or excessively difficult to exercise the rights provided by such provisions in practice. This arises by application of the principle of effectiveness, which is relevant in the present case on the following grounds. First, a large number of the claimants had been employed on a succession of fixed-term and temporary contracts and application of the six-month rule means that separate equal pay claims would have to have been brought in respect of each contract. Failure to instigate proceedings within the six-month time limit would disenable such individuals from advancing their claims. Secondly, the high degree of legal uncertainty that has surrounded the whole area of part-time workers' access to occupational pension schemes means that many would-be claimants were unaware that they had specific rights until, under the six-month rule, it was too late to do anything about it. There was, after all, an eight-year gap between the ECJ's ruling in *Bilka-Kaufhaus v Weber von Hartz*¹² in which the Court found that to deny part-time workers membership of an occupational pension scheme constituted indirect discrimination and the Court's ruling that such workers should be entitled to retrospective membership.¹³

Given such practical difficulties, it is hardly surprising that some of the claimants in the present case sought to have the usual six-month time limit

11 Case 33/76 *Rewe-Zentralfinanz Eg v Landwirtschaftsammer fur das Saarland* [1976] ECR 1989.

12 Case 170/84 [1986] IRLR 317.

13 See *Vroege* and *Fischer* cases cited in nn 1 and 2 above.

disapplied, arguing that its application in the circumstances would be contrary to the principle of effectiveness. When the case came before the Court of Appeal,¹⁴ it emerged that the central issue regarding the six-month limitation was not the imposition of the limit *per se*, but rather the date from which it should apply. The applicants argued that it should apply from May 1995 which was the date that a remedy was first provided under English law for the unlawful exclusion from such a pension scheme.¹⁵ The employers contended that the date of the ECJ's judgment in *Bilka*¹⁶ given in 1986 was appropriate. In the event, the Court of Appeal took the view that the absence of a specific legal remedy for an unidentified breach did not mean that the enforcement of Community law had been impossible. The House of Lords, however, felt that sufficient doubt existed to justify a reference to the ECJ. Given that much of the legal development in this context has unfolded in the European context, it is that Court's handling of the relevant issues which provides the most interesting interpretation of the law in this complex area.

It is noteworthy, that when the *Preston* case came before the ECJ the issues surrounding the application of the six-month time limit had shifted considerably between the Court of Appeal's ruling and the framing of the question referred by the House of Lords. The emphasis was no longer on when the six-month time limit should commence but, rather, on whether the provision of such a time limit could be maintained in line with the application of the effectiveness principle. This basic fact detracts from the importance of the ECJ's ruling as it will, presumably, be left to the House of Lords to decide on the real issue of impact: the date from which the six-month limit can reasonably be applied. This change in emphasis necessitates an exploration of the wider area of statutes of limitation and their treatment by the ECJ.

The temporal limitation in the *Barber* case

If the European Court's previous approach to the maintenance (or, to be more specific, imposition) of time limits is considered, certain salient issues begin to emerge. The application of limitations in retroactive claims is, of course, not a new development. In the *Barber* case,¹⁷ the ECJ clearly embarked on a departure from the general rule that its judgments have retrospective effect, even though the legal base of the claim derived from the fundamental right to equal pay provided by Article 119/EC. In restricting the right of new claimants to benefit retrospectively only from the date of the *Barber* decision, the Court recognised the need to balance the costs of implementing the judgment, (ie the resulting liability of those responsible for 'paying out') with the actual contributions received (ie the assets available from contributions 'paid in').¹⁸ In *Barber* the Court clearly accepted the argument that the necessity to achieve such a balance was fundamental to the successful administration of a pension scheme.

The Court justified its departure from the general rule on the grounds of the United Kingdom's reliance on the Occupational Social Security Directive 86/378/

¹⁴ [1997] ICR 899.

¹⁵ By the Occupational Pension Schemes (Equal Access to Membership) Amendment Regulations 1995 (SI 1995 No 1215).

¹⁶ See n 12 above.

¹⁷ See n 3 above.

¹⁸ See J. Moore, 'Justice doesn't mean a free lunch': the application of the principle of equal pay to occupational pension schemes' (1995) 20 ELR 159, 165-166.

EEC¹⁹ in the development of domestic policy relating to equal treatment in pension entitlement. The fact that the provisions of Directive 86/378/EEC were in conflict with the developing interpretation afforded to Article 119 only came to light in this context during the Court's consideration of the *Barber* case. The Court's inability to reconcile the provisions of the Directive with the fundamental right provided by Article 119 led to its finding that,

[t]he Member States and the parties concerned were reasonably entitled to consider that Article 119 did not apply to pensions paid under contracted out schemes and that derogations from the principle of equality between men and women were still permitted in that sphere.²⁰

As a result of the conflictual findings of the Court, Directive 86/378/EEC was amended²¹ and the member states took the unusual step of attaching a Protocol to the Treaty on European Union providing for retrospective effect only from the date of the judgment. This was, in itself, a surprising development which appears to have contradicted the ECJ's ruling in *Defrenne II* that the member states could not modify the legal obligations contained in Article 119 EEC by political declaration.²²

It should be noted that the Court's restrictive approach and the subsequent amendments to Community law were, in fact, the subject of controversy leading to widespread criticism despite the apparent justification for such action.²³ However, even if the Court's treatment of the *Barber* case is viewed as an acceptable attempt to reconcile the (conflicting) provisions of Community law with the potential costs of implementing the judgment, the temporal limitation imposed in *Barber* can be differentiated from the time bar relevant in the present case on two inter-related grounds. Firstly, the supremacy of European legislative provisions over conflicting domestic legislation should be considered in the context of the present case and, secondly, the impact of the jurisprudential development of the European Court of Justice on the evolution of domestic social policy is also worthy of exploration.

Supremacy of European law

The six-month limitation under consideration in the *Preston* claim derives, not from a provision of European law, but rather from a United Kingdom statute intended to give effect to Article 119 EEC. The Equal Pay Act 1970 may well predate the UK's accession to the European Community, but the requirement for domestic law to accord with the provisions of Community law has culminated in an amendment to the operation of the Act on a previous occasion²⁴ and the elimination of attempts to restrict the entitlement of claimants with respect to the amount of compensation available for equal pay and sex discrimination claims.²⁵

19 OJ 1986 L 225/40.

20 See the *Barber* judgment (1990) ICR 616, 259.

21 By Directive 96/97/EC.

22 In *Defrenne II*, the Court held that a Resolution adopted by the Member States could not modify the obligations owed under Article 119.

23 See n 18 above; E. Whiteford, 'Lost in the mists of time: the ECJ and occupational pensions' (1995) *Common Market Law Review* 801; S. Fredman, 'The poverty of equality: pensions and the ECJ' (1996) 25 *Industrial Law Journal* 91.

24 See Case 61/81 *EC Commission v United Kingdom* (1982) ECR 2601, in which the UK's lack of provision with respect to the facilitation of equal value claims led to the passing of the Equal Pay (Amendment) Regulations 1983 (SI 1983/1794).

25 The Sex Discrimination and Equal Pay (Remedies) Regulations 1993 (SI 1993/2798) removed the upper limits that previously applied.

In both the examples cited, the resulting alterations to national laws arose as a direct result of the jurisprudential development of the ECJ.²⁶

The developing jurisprudence of the ECJ

The second, and perhaps more compelling, ground on which the *Preston* claim can be differentiated from the *Barber* case relates directly to the relationship between European case law and the development of social policy within member states. In the *Barber* case, the Court was confronted with a hitherto unencountered application of European legislation which, when considered fully, resulted in the Court finding that certain provisions of Directive 86/378 were in direct conflict with Article 119 EEC. As we have seen, this ruling led to the amendment of the Directive and the striking of a 'compromise' in the implementation of the judgment between the needs of UK employers and pensions administrators and employees by way of the temporal limitation. It would appear, however, that the parties in the present case should have been assisted by a more certain legal environment due to the developments that have taken place in the case law context in the decade since *Barber* was considered by the ECJ. During this time there have been significant developments in the application of the equality principle in the context of both occupational pensions and part-time employment with corresponding overlap. The development of social policy within the Member States should surely have kept apace with such advances, particularly given the wide publicity that the *Barber* judgment received.²⁷

The member states have known that the equality principle applies, without exception, to occupational pension schemes at least since the ECJ's ruling in *Barber* in 1990 and, in the context of part-time workers' access, since the Court's 1986 decision in *Bilka-Kaufhaus*. That a failure to act on such information should now be sanctioned by the Court by an application of the six-month time limit imposed by domestic legislation appears to let such tardiness off the hook twice over. Conversely, the claimants denied access to legal redress due to the fact that the provisions of domestic law did not offer any legal basis on which a claim could be advanced, have, if the six-month time limit is maintained, been doubly thwarted. Employers, on the other hand, may now find themselves in the unenviable position of having to meet costs that they did not expect to incur.

Accordingly, it could be contended that the UK's failure to amend the provisions of domestic law in this respect amounts to a failure to adequately implement European social law. This argument was advanced unsuccessfully when the *Preston* case came before the Court of Appeal. On that occasion, the Court found that the applicants were unable to rely on the ECJ's ruling in *Emmot v Ministry of Social Welfare*²⁸ that a state is not able to rely on its own failure to implement the requirements of Community law in contending that a claim is time-barred. In such

26 In Case C-271/91 *Marshall v Southampton and South West Hampshire Health Authority (No 2)* (1993) ICR 893, the ECJ held that the imposition of an upper limit in such cases was in contravention of Article 6 of the Equal Treatment Directive 76/207/EC.

27 Domestic legislation on pensions was in fact substantially amended following the Court's decision in *Barber* by the Pension Schemes Act 1993 and, following the decision in Case C-200/91 *Coloroll Pension Trustees Ltd v James Richard Russell and Others* [1994] ECR I-4389, by the Occupational Pension Schemes (Equal Access to Membership) Amendment Regulations 1995 (SI 1995 No 1215) which provide, *inter alia*, that the full cost of back-dated pension benefits in respect of periods of service after 31 May 1995 should fall to the employer.

28 Case C-208/90 [1991] ECR I-4269.

cases, the time limit provided under domestic law would only commence once adequate implementation had been achieved. However, this is only relevant in cases involving failure to implement a directive, not a Treaty Article, and only applies when the state is defendant. Both of these conditions preclude the application of the Court's common-sense ruling in *Emmot* to the present case.

Although it was not directly considered by the ECJ, the wider question of what date to apply the six-month limitation from may well resume its centre-stage position when the case is reconsidered by the House of Lords. The employers will, presumably, hope for an extension of the relevant period from as far back as possible in order to maximise the number of claims that will be time-barred.²⁹ The claimants, on the other hand, will seek confirmation that an appropriate degree of legal certainty was only achieved by the introduction of a remedy in 1995, thus potentially enabling the advancement of a larger number of claims that were not instigated timeously. However, given the decision in the *Emmot* case and the guidance provided by the ECJ in its ruling on *Preston*, it seems likely that the House of Lords will adopt a strict interpretation of the provisions of the EqPA which allow only those proceedings instigated during, or within six months of the cessation of the period of employment to be advanced.

The principle of equivalence

Whatever yardstick is selected by the House of Lords, the future operation of the law will almost certainly rest on an interpretation of the relevant developments of European case law. In *Levez*,³⁰ the ECJ ruled that it is for the national courts to decide whether procedural rules or other conditions which provide an alternative remedy under domestic law are more favourable than those which implement Community law. Such provision would be prohibited by the principle of equivalence which deems that the procedural rules for safeguarding rights derived from Community law should not be less favourable than those governing similar domestic actions.

In the present case, the fact that the procedural rules relating to breach of contract and tortious duty permit a period of up to six years to elapse between the breach and the instigation of proceedings,³¹ may offer a window of opportunity for extension of the six-month limit. However, the ECJ, when giving its ruling in *Levez*, drew a clear distinction between the rights emanating from an application of the provisions of the Equal Pay Act 1970 and rights provided by other domestic provisions. It was up to the national courts to determine whether the purpose and essential characteristics of other domestic actions were sufficiently similar so as to warrant application of the principle of equivalence.³² Unless such similarity can be established, even where more favourable time limits operate under the application of domestic rules, member states are not required to extend their most favourable rules to all actions brought in the field of employment law.³³

29 As discussed above, the date of the ECJ's ruling in *Bilka* was advanced as the appropriate date before the Court of Appeal. If the House of Lords were to accept this as the date from which claimants should have instigated proceedings, those who failed to do so in the nine years between the decision in *Bilka* and the introduction of the Occupational Pension Schemes (Equal Access to Membership) Amendment Regulations 1995 would find their claims time-barred.

30 See n 6 above.

31 By virtue of the provisions of the Limitation Act 1980.

32 See *Levez* cited in n 6 above, para 43.

33 *ibid* para 42.

It is interesting to note that the usual time limit imposed by UK statutes concerned with the provision of rights in the employment context is actually three months,³⁴ with employment tribunals granted discretion to extend such limits on the grounds of reasonable practicability. The ECJ has held time limits provided by domestic legislation in other member states, which are much shorter than the six-month limit provided by the EqPA, to be compatible with Community law as long as they are deemed to be reasonable in the circumstances.³⁵

Claims brought within the time limit

Many of those whose claims are not time-barred may find the practicalities involved in advancing them rather more complex than at first apparent. This is because the right to apply retroactive membership to contributory schemes will only be granted if those claiming such membership make payment of the appropriate contributions as a lump sum. This was established by the ECJ in the *Fischer* judgment, in which the Court held,

[T]he fact that a worker can claim retroactively to join an occupational pension scheme does not allow the worker to avoid paying the contributions relating to the period of membership concerned.³⁶

In practical terms this may present specific difficulties for the claimants involved, as the denial of entitlement to an occupational pension will have left many of the affected individuals reliant on partners' occupational pensions or state entitlements. Given the passage of time relevant to some of the claims it is, thus, unlikely that many of the women involved will have access to the necessary funds.

In addition, the moral dimension of the operation of the law following the ECJ's treatment of this whole issue raises further questions. Firstly, the financial hardship, which may well prevent some of the would-be claimants from accessing the pensions to which they are legally entitled, will undoubtedly benefit the employers involved. This is because, although the employers would be obliged to match any contributions forthcoming from the claimants in order to create a fund from which pensions would be paid, any potential sums which the women are unable to pay will not require to be met. In this respect it appears that the financial cost to employers has been minimised at the expense of the pension entitlement of these 'successful' claimants. Furthermore, many of the original claimants have died during the course of this protracted litigation. The lack of the necessary salary information is likely to prevent surviving spouses and other would-be beneficiaries from being able to prove entitlement.

Finally, the value of the pensions received by those who are able to make the necessary financial contributions bears some consideration. Even where the funds are forthcoming from both employer and former-employee, it is unlikely that the actual sum received will be equal to that which would have been available had the worker not been excluded from membership in the first place. Pension funds are

34 For example, the right to claim unfair dismissal provided by Part X, Employment Rights Act 1996; the rights emanating from the Sex Discrimination Act 1975; the Race Relations Act 1976 and the Disability Discrimination Act 1995 and the rights connected with trade union activities and duties provided by the Trade Union and Labour Relations Consolidation Act 1992.

35 For example, the time limit applicable in the German case *Rewe* ([1976] ECR 1989) was one month; in the Dutch case, *Comet* ([1976] ECR 2043), 30 days was held to be a reasonable limit. See Opinion of Advocate-General Léger in the present case.

36 Case C-128/93 [1994] ECR I-4583, 665.

administered by the investment of contributions in order to generate future income which is then used to pay out benefits. It is unclear how the current value of the claimants' entitlement is to be assessed and who will bear the cost of the interest that would have accumulated had the sums actually been invested.

Notwithstanding the potential hurdles that claimants are likely to face, the cost for employers in certain sectors may well be considerable. The Institute of Directors has estimated a £17 billion price tag³⁷ and public sector employers, in particular, are likely to be greatly affected due to the large numbers of female part-time workers within that sector of the workforce. To date, the Government has declined to issue any statement regarding the potential costs of implementing the judgment.³⁸ However, the National Union of Teachers has estimated that a typical teacher with service amounting to 25 years, of which 12 years was undertaken on a part-time basis, and who retired on a salary of £24,000 would be eligible for an extra £1,800 per year in pension entitlement plus a lump sum of £5,400.³⁹

It should be remembered that the high costs now faced by employers have arisen due to the fact that so many workers were denied entitlement to occupational pensions in the first place. This omission and its subsequent defence by policy-makers encapsulated the UK government's past attitude to part-time workers. The historical treatment of such workers as second-class citizens, with reduced entitlement to benefits and employment protection in comparison to their full-time counterparts, may well have been consigned to the past given recent common law developments and the implementation in the UK of the Part-time Work Directive 97/81/EC.⁴⁰ However, it seems likely that at least some of the old inequality will remain as evidenced by the restrictive approach adopted by the Part-time Workers (Prevention of Less Favourable Treatment) Regulations 2000.⁴¹ The rights afforded to part-time workers by virtue of the implementing legislation will depend on such a narrow range of comparison that those who lack a full-time worker within the same employment with whom to compare themselves will be denied any protection against discrimination on the grounds of their part-time status. The apparent lack of compliance with the intended spirit of the European legislation can only lead one to question whether any lessons have been learned from the past, particularly given the case under consideration.

Conclusions

The ECJ's ruling in the *Preston* case poses more questions than it answers. It will fall to the House of Lords to decide on the enforceability of the six-month time limit and, if it is maintained, the date from which it should apply. In giving its judgment, the Court will have to achieve a careful balancing act. On one hand, consideration will have to be paid to the needs of employers who may be required

37 This figure has been disputed by other sources, for example, actuaries William Mercer calculate total costs for employers at £100m – see *Scotland on Sunday*, 21 May 2000, 'Jumping through hoops in wacky world of pensions' 16. However, figures stating the actual bases of such estimates have proved elusive.

38 In a recent written answer, the Secretary of State for Social Security gave the following response to a question regarding the costs involved, 'This is a complex ruling which needs careful consideration' and resisted commenting further until the case has been reconsidered by the Lords. See House of Commons Hansard, written answers for 23 May 2000, Column 484W.

39 From the *Guardian*, 17 May 2000, 'Pension ruling will cost employers dear' 4.

40 Made 15 December 1997, OJ L014, 20 January 1998, 9–14.

41 SI 2000/1551, made 8 June 2000, came into force on 1 July 2000.

to meet costs that they have not prepared for. On the other hand, the claimants will be looking for an acceptable conclusion to a long and hard-fought struggle for recognition of what is, *prima facie*, their legal entitlement. The fact that the final conciliation of these opposing positions will rest on the Lords' interpretation of a legal technicality will offer little comfort to either side. The distinction between the limitations in the periods of service that may count and the application of time limits, although crucial in this case, makes no sense to laymen, whether employers or employees.

All this must be considered against the backdrop of a domestic legal system in which progress to date has been frustratingly incremental. In seeking to ensure that the relevant domestic provisions accord with the principle of equivalence required by Community law, the stubborn refusal of the UK to act on the guidance proffered by the jurisprudential treatment of related claims since as far back as *Bilka* will inevitably loom large. The manner by which their Lordships decide to reconcile this aspect in the context of the present case remains to be seen, but one factor seems likely. Due to the legacy of the past, the employers' costs will be minimised at the expense of the remedies available to those part-time workers awaiting clarification.

REVIEW ARTICLE

The Theoretical Controversy Concerning Judicial Review

*Andrew Halpin**

Christopher Forsyth (ed), Judicial Review and the Constitution, Oxford: Hart Publishing, 2000, xxxvii + 442 pp, hb £35.00.

The controversy concerning the theoretical basis for judicial review is a matter of importance that extends beyond the particular concerns of public lawyers. Embracing the fundamental relationship between the legislature and the judiciary and their respective constitutional roles, it brings into sharp focus issues which have a bearing on our wider understanding of the nature of law and its practice. The way the controversy has developed provides a particularly striking opportunity to consider the benefits of relating academic and practitioner perspectives, and to consider the value of theoretical insights for our understanding of the practice of law. The course this particular controversy has taken also raises general questions about the nature of academic debate, and the way that academic and practitioner alike are inclined to represent the law. Accordingly, a wide variety of readers should welcome this collection of essays on the foundations of judicial review.¹

The two models of judicial review

Two models have been proposed for the basis of judicial review. A certain amount of refinement has occurred to each, but it is helpful to start with the extreme versions so as to clarify matters before undertaking the more difficult task of exposing the real nature of the dispute that now remains.²

The *ultra vires* model treats the basis for judicial review as the courts ensuring that a power granted by the sovereign Parliament is not exercised beyond the limits

* Faculty of Law, University of Southampton. I am grateful to the AHRB for an award to support a research project of which this essay forms a part, and to Liz Fisher for helpful discussions of judicial review.

1 Unattributed references below are to this collection.

2 It is not altogether clear that either model in its extreme form has gained unalloyed support. The extreme form of the *ultra vires* model is associated with Sir William Wade (Mark Elliott, 84). The extreme form of the common law model is thought to have surfaced in what Lord Irvine has described as some 'extra-judicial romanticism' of Lord Woolf and Sir John Laws (referred to at 45, 126–127, 394). Reasons why it is difficult to find a pure expression of either model in its extreme form include the varied range of contexts in which judicial review is sought, the confusing range of uses to which the term *ultra vires* has been put (on which see David Feldman, ch 11), and the speculative nature of some of the discussion. Nevertheless, even if the two models in their extreme forms represent hypothetical theoretical positions, they have their value in marking out the nature of the gap which has been narrowed in subsequent theoretical positions actually adopted.

which Parliament sets. The principles of judicial review are, accordingly, just those which mark out the limits found in the explicit or implicit intention of Parliament in granting the power.

The common law model (again in its extreme form) treats the principles of judicial review as created by the common law, judicially constructed to ensure the proper exercise of power. As such they could even be used to strike down a power explicitly granted by Parliament to be exercised in breach of these principles.

It is apparent that the contested ground between these two models in their extreme forms covers the sovereignty of Parliament and judicial creativity. The modified *ultra vires* model still maintains the role of the courts as subordinate to the sovereign Parliament but now acknowledges a creative judicial role in working out the detailed principles of judicial review. These roles of Parliament and the courts are kept in harmony by positing an implied legislative intent that all statutory powers should be exercised in accordance with the principles of judicial review which Parliament has authorised the courts to apply. The modified common law model openly acknowledges the sovereignty of Parliament by admitting that the principles of judicial review could not be effective against a clear statutory statement that they should not prevail, but still maintains an independent common law source for the principles, which are regarded as having come into being irrespective of Parliamentary intention.

What is apparent now is the common ground between the two models. Although impartial contributors to the debate find it less easy to see what the remaining differences are, in the eyes of the leading proponents of the two models the distinction is still very real. In considering what remains of this distinction I shall concentrate on the features that make the two models what they are. Such an exercise should help us to unravel the unresolved issues at the heart of judicial review.

The 'undistributed' or 'excluded middle'

One issue affecting the core features of the two models surfaces at various stages in the debate, and is alluded to by both Paul Craig and Christopher Forsyth as still very much a part of what divides the two approaches. Craig, in his assessment of the remaining difference between the two models, states that:

It should be remembered that the central kernel of the argument [for the *ultra vires* model] is that where Parliament does not prohibit it must be taken to authorise expressly or impliedly. It is for this reason that legislative intent must be found in order to legitimate the imposition of any constraints on the way in which power is exercised. ... The common law model holds that the principles of judicial review are developed by the courts. ... If the omnipotent Parliament does not like these controls then it is open to it to make this explicitly clear. ... There is therefore nothing in the common law model which involves a strong challenge to sovereignty. (pp 378, 382-383)

The contrasting perspective is provided by Forsyth:

The difference between [the two models] is simply over the articulation of what is plainly an artificial construct: the intention of Parliament. Should Parliament be presumed to have authorised the application by the judges of the principles of good administration or should that authorisation come from the common law? ... The concepts of *ultra vires* and *intra vires* are mutually exclusive: a decision-maker either acts within or outside his or her powers, there is no middle ground. ... If, say, a decision-maker in denying a hearing in certain circumstances, acts within the powers granted by Parliament the common law cannot impose

a duty of fairness upon that decision-maker without challenging parliament's power to allow him or her to make valid decisions without a hearing. Otherwise you have the position where every requirement for validity laid down expressly or implied by Parliament is satisfied yet the common law is imposing an additional requirement for validity. That is a challenge to parliamentary supremacy. (pp 396, 402)

The issue arising in these extracts is whether the intention of Parliament covers the conditions under which the power that it has granted is to be exercised, or whether it leaves a gap for the courts to impose conditions. This is the issue of the so called undistributed or excluded middle.

The logical law of the excluded middle asserts that something must either be or not be the case ($p \vee \sim p$). There is no room for a situation in the middle of these two alternatives. Forsyth calls upon the notion (though not the terminology) of an excluded middle in the following way:

The analytical difficulty is this: what an all powerful Parliament does not prohibit, it must authorise either expressly or impliedly. Likewise if Parliament grants a power to a minister, that minister either acts within those powers or outside those powers. There is no grey area between authorisation and prohibition or between empowerment and the denial of power. (pp 39–40)

Sir John Laws, on the other hand, argues that there can indeed be a case in between 'Parliament authorises' and 'Parliament prohibits':

Forsyth's argument is vitiated by an implicit mistake: the mistake of assuming that because Parliament can authorise or prohibit anything, all authorities and prohibitions must come from Parliament. It is a *non sequitur*. It neglects what the logicians call the 'undistributed middle' – an obscure, but useful, academic expression, meaning that although X and Y may be opposites, like praise or blame, they do not cover the whole field; there might be Z, which involves neither. (p 78)

The terminology is misplaced for the principle of the undistributed middle technically refers not to the middle that was excluded by $p \vee \sim p$, but to the middle term in a syllogism.³ Nevertheless, the gist of Laws' argument is clear. He is denying that we have a case of an excluded middle. Sometimes it is the case that Parliament neither authorises nor prohibits.

In denying that there exists on a particular occasion a case of the excluded middle, one is not necessarily denying the law of the excluded middle. One may simply be pointing out that the case of suggested alternatives does not fall under $p \vee \sim p$ because the one alternative is not the negation of the other, and hence the law of the excluded middle is not relevant. This is certainly the position with the analogy that Laws draws with praise or blame. Blame is not formed by the negation of praise – from the fact that I do not praise you it does not mean that I blame you. So on this occasion we do not have a case of $p \vee \sim p$ to fall under the general law.

The analogy used by Forsyth, pregnant or not pregnant, is a perfect illustration of the law of the excluded middle – we reach the one alternative by the negation of the other, and $p \vee \sim p$ holds. The crucial question is: what is the relationship between the alternatives of Parliament intending to attach a condition to a power it has granted, and Parliament not so intending?

3 For example, an argument to prove that all sharks are lawyers:

All lawyers can bite;

all sharks can bite;

therefore, all sharks are lawyers.

is flawed because the middle term, 'can bite', is undistributed (ie does not cover all of its instances).

Let us drop the technical terminology for the moment. It is perfectly intelligible to say of a person, for example, in relation to seeing a particular film at the cinema this evening, that he:

- (1) wants something to happen, or
- (2) wants something not to happen, or
- (3) has no wishes one way or the other

This does not breach the law of the excluded middle, for (1) and (2) are separate propositions, each of which with its negation can satisfy the law of the excluded middle. Recognition of the distinct negations of (1) and (2) is a prerequisite to the recognition of (3), which is formed by the conjunction of the negations of (1) and (2).

We can similarly provide a fuller presentation of the range of possibilities when Parliament grants a power, it is possible, for example, that:

- (1a) P intends the power to be exercised reasonably, or
- (1b) P does not intend the power to be exercised reasonably.
- (2a) P intends that the power should be exercised unreasonably, or
- (2b) P does not intend that the power should be exercised unreasonably.

and where both (1b) and (2b) hold

- (3) P has no intention whether the power should be exercised reasonably or not.

The logical possibility of (3) becomes more plausible as a practical proposition when there has been no opportunity for consideration of what would be regarded as reasonable in relation to a specific exercise of the power.

‘An all-powerful Parliament’

Establishing that the law of the excluded middle does not make it logically impossible for Parliament to have no intention one way or the other, in relation to placing a condition on the exercise of a power it has granted, does not end the debate. There may be further reasons why Parliament should be regarded as having closed down all possibilities for the making of conditions other than those it has itself authorised.

It is clear that for Forsyth, and the other supporters of the *ultra vires* model, the mainstay of their argument is the threat to parliamentary sovereignty that they believe would follow were their model abandoned in favour of the common law model.⁴ Forsyth, in his concluding contribution, regards it as the principal strength of the *ultra vires* model that it ‘provides a firm constitutional foundation for judicial review’ by treating the principles of judicial review as not a threat to parliamentary sovereignty but as impliedly authorised by Parliament (pp 408, 404). Mark Elliott stands on the same ground, and has issued the ultimate challenge to the other side:

so long as the common law accords a legislative supremacy to Parliament, it must be possible to reconcile the courts’ public law jurisprudence with this constitutional principle. It is the interpretive methodology of *ultra vires* – and *only* this methodology – which is capable of securing this reconciliation. (p 109)

⁴ Mark Elliott is a particularly strong ally on this point – see his chs 5, 12 and 16, and ‘The Demise of Parliamentary Sovereignty? The Implications for Justifying Judicial Review’ (1999) 115 LQR 119.

This challenge assumes certain things about the nature of parliamentary sovereignty. Given that Parliament could be in a position of indifference as to the imposition of a condition on a power it had granted, the argument on the side of the *ultra vires* model has to place all its weight on an understanding of parliamentary sovereignty capable of excluding the logical possibility conveyed by proposition (3). In examining this further we will need both to consider the understanding of parliamentary sovereignty involved, and also to look in more detail at the practical factors which might bring about a case falling under proposition (3).

In relation to parliamentary intention concerning a matter capable of being governed by law, it may be the case that Parliament has considered the matter and decided upon a neutral stance towards it.⁵ It may be that there has never been an occasion when Parliament has debated anything even remotely connected to the matter, or that the matter was not thought about at the time when a related subject was being debated. It may be that although there was general awareness of this particular matter, it was not brought to Parliament's attention or even implicitly taken into account when related legislation was passed. It may be that although the matter was raised in Parliament, the timetable of parliamentary business precluded relevant parliamentary intention being formed before the related legislation was enacted. It might even be the case that it was expedient not to resolve a parliamentary intention in relation to that particular matter before the vote was taken.⁶

Parliamentary sovereignty is a subject which can provide theorists with a number of stimulating issues to pursue.⁷ It may be as much an artificial construction as parliamentary intention,⁸ but even so may perform the function of covering, albeit figuratively, doctrines of practical importance. For our present interests we may fortunately reduce the subject matter of parliamentary sovereignty to a small number of practical concerns.

It is one thing to invoke the doctrine of parliamentary sovereignty in proclaiming (in the anthropomorphic language so often used in relation to Parliament) that Parliament cannot bind its successors, or in asserting that what Parliament enacts is supreme. It is, however, something else entirely to claim that an aspect of parliamentary sovereignty requires us to conclude that nobody can do anything without Parliament authorising it. This at the very least turns Parliament into an

5 The objection that Parliament by adopting a position of indifference is in fact adopting an intention in favour of the status quo, and so can never be truly indifferent, assumes that there exists a legal status quo. This is not always the case. It is not the case where the existing law is unclear and Parliament adopts a position of indifference as to its future course by declining the opportunity to enact a law to clarify it one way or the other (some would see the failure to codify the criminal law as a matter of parliamentary indifference). Nor is it the case where a change in the law is brought about with uncertain consequences and Parliament adopts a position of indifference as to which of those future consequences will prevail by declining the opportunity to deal with the matter in the legislation (whether in leaving it to the courts, or in leaving it to a statutory instrument to fill in the details).

6 The skill of the draftsman is occasionally called upon to word legislation that is ambiguous as to the determination of such a matter, so that every side within the apparent parliamentary consensus can be appeased. Louis Jaffe, *Judicial Control of Administrative Action* (Boston, MA: Little, Brown & Co, 1965) 37–38, suggests that the inability to reach a detailed consensus within the legislature on controversial matters may be a good reason for delegating power to an administrative body.

7 For some recent examples, see Jeffrey Goldsworthy, *The Sovereignty of Parliament: History and Philosophy* (Oxford: Clarendon Press, 1999); Neil MacCormick, *Questioning Sovereignty: Law, State, and Nation in the European Commonwealth* (Oxford: Oxford University Press, 1999); N. W. Barber, 'Sovereignty Re-examined: The Courts, Parliament, and Statutes' (2000) 20 OJLS 131; Paul Craig, 'Public Law, Political Theory and Legal Theory' [2000] PL 211, 211–230; and Martin Loughlin, *Sword and Scales: An Examination of the Relationship between Law and Politics* (Oxford: Hart Publishing, 2000) 136–140, 151–157.

8 It was once regarded as 'almost entirely the work of Oxford men'. See R.F.V. Houston, *Essays in Constitutional Law* (London: Stevens & Sons, 2nd ed, 1964) 1.

interfering busybody. In fact, it requires of Parliament super-human capacities even in its anthropomorphic guise. It would mean that Parliament had the capacity to consider each past circumstance and each future eventuality, and to formulate an appropriate authorisation to deal with it. It would require the capacity to eradicate all of the cases that we have just catalogued as leading to a recognition of proposition (3). This needs more than an 'all-powerful Parliament'. We would need a legislative body displaying omniscience and capable of acting beyond temporal constraints.⁹

Recognising that Parliament has neither the time nor the capacity to exercise its sovereignty in such an officious manner does not lead us to deny other aspects of its sovereignty. The common law model adopts a practical view of parliamentary sovereignty which recognises that nobody can tell Parliament what to do, and that Parliament can always have the final say, but considers that until Parliament sees fit to proclaim otherwise the courts will get on with their business of formulating and applying the principles of judicial review. 'If the omnipotent Parliament does not like it, it can say so.' (To paraphrase slightly Craig's retort.)

Admittedly, this view of parliamentary sovereignty acknowledges that the courts continue to go about their business on the sufferance of Parliament: that the sovereign Parliament if it wished could take away all the business of the courts, and hand it over to popular tribunals; and could undo all the past work of the courts by enacting a comprehensive code to replace all existing law. Still, this view does not recognise that the conduct of each element of the court's business is authorised by Parliament. It is rather like, in a more primitive context, trying to account for the continuing validity of the laws of Rex I after his death upon the succession of Rex II. Given that Rex II could have repealed all the laws of Rex I but has not done so, their continuing validity is on the sufferance of Rex II. However, it would be absurd to suggest that Rex II authorised the laws that Rex I made.¹⁰

It is part of the common ground between the modified *ultra vires* and common law models that the details of the principles of judicial review are created by the courts, so neither side is suggesting that in some way parliamentary sovereignty requires Parliament to authorise the material content of the principles of judicial review. Yet there remains a point of issue between the two models over the matter of who has authorised the principles of judicial review. Are they created under the authority of Parliament, or by the courts' own authority?

What exactly are the practical implications of the view of parliamentary sovereignty required by the *ultra vires* model? Since it is not being suggested that Parliament authorises the material content of the principles of judicial review, how exactly does its purported authorisation of the principles bite in a practical way? An important qualification that has been suggested from the *ultra vires* side is that the issue of authorisation only arises in the context of the courts dealing with the conditions that are to be attached to a *statutory* power.¹¹ This recognises that the courts on their own authority get on with the business of developing and applying the principles of judicial review for non-statutory bodies, but insists that any conditions attached to a power granted by Parliament must be authorised by Parliament itself at pain of threatening parliamentary sovereignty.

⁹ It is not sufficient to rely on the default intention enthusiastically adopted by John Austin in W. Rumble (ed), *The Province of Jurisprudence Determined* (Cambridge: Cambridge University Press, 1995) 35. This is either fiction, suggested by Austin's remark that rules made by judges are 'established by the sovereign legislature'; or another instance of the error of equating the absence of a prohibition with a positive authorisation, suggested by his talking of an authority imparted 'by way of acquiescence'.

¹⁰ See H.L.A. Hart, *The Concept of Law* (Oxford: Clarendon Press, 1961) 60–64.

¹¹ See Elliott's discussion of this point at pp 90–91

However, the recognition of proposition (3) and the variety of ways in which a case may fall under it, means that there is no direct challenge to the will of Parliament until Parliament has had and exercised the opportunity to express an intention upon each of the conditions that might or might not be regarded as an appropriate limitation of the power. On a smaller scale, the super-human qualities required by the doctrine of parliamentary sovereignty in its hyper-officious form would be needed here to maintain the position that all the conditions attached to a power granted by Parliament must be authorised by Parliament itself. Yet it is implicit in the acknowledgment by the *ultra vires* side that the material content of the principles of judicial review are the creation of the courts that this is not presumed.

One way out of the apparent lack of an effective notion of parliamentary sovereignty on the *ultra vires* side is to resort to a romantic view of parliamentary sovereignty. It is, strictly speaking, unnecessary to be conscious of parliamentary sovereignty when the courts are working out the practical determination of the principles of judicial review, but to restrict our perspective in this way misses out on the beauty of the broader picture. The courts are not simply doing their own thing but working in harmony with principles that Parliament also embraces and fully endorses. Something like this seems to be the impetus for Elliott's heartfelt request:

In order to acknowledge that the credit lies with the courts for the translation of these norms into enforceable legal principles, it is not necessary to deny that Parliament has long been taken to respect those values. There is no shame in admitting that judicial development of administrative law is consistent with the legislative intention of Parliament. ... [I]t is, rather, the characteristic of a mature democracy in which respect for the most fundamental values permeates each branch of the constitution. (pp 368–369)

If it were known that a sovereign Parliament endorsed certain values, then there could be something in an argument that suggested the courts were offering an affront to parliamentary sovereignty by adopting any other values in applying principles of judicial review. However, this takes the *ultra vires* model onto much weaker ground.

A return to the realm of fairy tales

Reference is made to two sorts of values by the supporters of the modified *ultra vires* model, which it is confidently assumed can be regarded as governed by parliamentary intention: the standards of reasonableness, and the values of the rule of law. Essentially the same point can be made about both in revealing the flimsiness of their romantic appeal. They both have an intangible quality when it comes to the point of capturing them in concrete situations. I shall concentrate on reasonableness here, reserving comments on the rule of law for the following section.¹²

The question whether a power has been exercised reasonably is usually raised by a victim of the exercise of that power, who wishes to challenge it. In court the standard defence in such an action for judicial review is not that the body exercising the power enjoyed the authority to exercise it unreasonably, but that the power had been exercised reasonably. This alone should be sufficient to alert us to

12 Elliott does suggest that reasonableness is one of the requirements of the rule of law (96), but this view is not necessarily a part of the argument whenever it is asserted on the *ultra vires* side that the requirement of reasonableness must be regarded as being governed by parliamentary intention.

the fact that the standards of reasonableness are not settled. Moreover, even if we narrow it down to reasonableness in the context of what purposes the power was granted to achieve, the points of contention simply multiply. What exactly are the purposes? How far is it envisaged that those purposes should be achieved at the expense of competing considerations?

If uniform standards of reasonableness existed, it would not matter whether we took those standards as internal to the grant of power or externally imposed. The question whether Parliament intended the power to be exercised reasonably or the courts had imposed the condition would be redundant for all practical purposes (and probably for all theoretical purposes too) if there existed a uniform set of standards of reasonableness known to all. For when Parliament granted the power it would know exactly under what conditions, subject to that uniform set of standards, the grant would take place. And when the courts reviewed the exercise of the power they would simply apply that uniform set of standards to the case before them (and presumably litigation in this area would dry up to leave only disputes over the facts).

Only a cynic would suggest that every human endeavour to behave reasonably is a transparent illusion. Nor would it be so fanciful to hold a general presumption that people are capable of behaving reasonably, and in a trivial sense retain the expectation that all will behave reasonably. However, what is being imputed to Parliament is not a lofty view of mankind but an authorisation of a set of standards of reasonableness used for the principles of judicial review, by which hard decisions are made as to which exercise of power was reasonable and which was not. It does not require a cynic to point out that no human effort has so far achieved an intelligible set of standards of reasonableness to determine what amounts to reasonable behaviour in any given situation. It is nothing more than a fairy story to suggest that Parliament has worked out and adopted a set of standards of reasonableness, and a bigger story to suggest that Parliament shares the same set of standards with the courts.¹³ Breaking the spell of these illusions leads to a number of conclusions.

First, the absence of a settled set of standards of reasonableness means that there exists the possibility of dispute over what precise standards should be applied, how they should be prioritised, and how they should affect the outcome, in a particular concrete situation. Secondly, it is worth repeating that without a settled body of standards (unless Parliament works through all practical eventualities and forms its own view upon them), the absence of an intention that a power should be exercised reasonably should not lead us to infer that Parliament was consciously indifferent to the power being exercised unreasonably. Thirdly, we may properly characterise reasonableness as encompassing standards to be worked out before they can be applied, as involving a deliberative endeavour rather than a communicative requirement.

Replacing fig leaves

The criticism that the reliance on an implied Parliamentary intention was nothing more than a fig leaf used to cover up the courts' own creative role, in producing

13 I leave aside differences between the views of reasonableness of one judge and another. For fierce exchanges on this, see Richard Revesz, 'Environmental Regulation, Ideology, and the DC Circuit' (1997) 83 *Virginia Law Review* 1717; Harry Edwards, 'Collegiality and Decision Making on the DC Circuit' (1998) 84 *Virginia Law Review* 1335; Richard Revesz, 'Ideology, Collegiality, and the DC Circuit. A Reply to Chief Judge Harry T. Edwards' (1999) 85 *Virginia Law Review* 805.

principles of judicial review to set the conditions to the powers that Parliament had granted, is less stinging in the light of the modifications that the two theoretical models of judicial review have undergone. Both now acknowledge a creative role for the judiciary. As we have seen, the point of issue remaining between the two models is about who authorises the principles of judicial review that the courts apply with their creative ingenuity to the situations arising before them. Both sides also have described the principles, that each from its own perspective gives different credit for authorising, as amounting to the values of the rule of law. As I indicated in the previous section, it would I think be possible to take the values of the rule of law through the same argument that was employed there on the standards of reasonableness. The result would be to show that the claim by the *ultra vires* side that Parliament must be taken to have authorised the values of the rule of law as used by the courts in judicial review is as much a fairy story as the claim in relation to the standards of reasonableness.¹⁴ However, I think that there is a more worthwhile exercise to be undertaken in considering how the rule of law has been used in this debate. This arises out of the observation that the use of the rule of law differs between the two models. If its use as a prop in a fairy story falls to the *ultra vires* side, I shall suggest its use as a fig leaf is now more evident on the common law side. I shall actually suggest that the common law side is outdoing the *ultra vires* side by collecting a wardrobe of fig leaves.

On the *ultra vires* side the use of the rule of law (as with reasonableness) is as an instrument to connect Parliament to the principles of judicial review applied by the courts. The fact that this instrument happens to be a set of values is secondary to the fact that it is regarded as being held in common by both, so that Parliament can be taken to have implicitly authorised what the courts engage in creatively applying. I shall rely on the suggestion made above that the argument developed in the previous section can meet this particular contention. On the common law side, by contrast, the rule of law is used precisely because it amounts to a set of values. This set of values provides an authority for the business of the courts. The different uses of the rule of law relate directly to the outstanding point at issue between the two models. The *ultra vires* side sees the sovereignty of Parliament as providing the authority for the principles of judicial review so has no need to rely on values in the rule of law to provide authority. The common law side in denying that the authority comes from Parliament, and wishing to divert our gaze from the possibility that the principles are nothing but bare judicial constructs, has recourse to the values seen as inherent in the rule of law to provide authority for the principles that the courts employ. The old fig leaf of parliamentary intention has been replaced by a new fig leaf of the rule of law.

This distinction between the two uses of the rule of law is apparent in the following passages. For the *ultra vires* side Elliott states:

It is the simple – and wholly plausible – assumption that Parliament intends to legislate consistently with the rule of law which bridges the apparent gulf between legislative silence and the developed body of administrative law ... (p 96, emphasis added)

On the common law side, however, Craig states:

14 That the values (or even the doctrine) of the rule of law are not settled appears from the contributions to this volume by David Dyzenhaus (ch 7) and T.R.S. Allan (Comment). For further discussion, see a particularly pertinent contribution to David Dyzenhaus (ed), *Recrafting the Rule of Law: The Limits of Legal Order* (Oxford: Hart Publishing, 1999) by Alon Harel, 'The Rule of Law and Judicial Review: Reflections on the Israeli Constitutional Revolution'.

Supporters of the common law model have made it clear that they believe that the content of judicial review should be decided in accordance with the rule of law, or normative considerations of justice, *which warrant the imposition* of constraints on the exercise of discretion. (p 376, emphasis added)

This purported derivation of authority from the rule of law presumes, as much as did the purported derivation of authority from parliamentary intention on the *ultra vires* side, a settled body of values through which the authority can be transmitted. Once we acknowledge the contentious nature of the rule of law, no such settled body of values can be found. At one point, Craig himself acknowledges 'the different meanings attached to the concept of the rule of law' as grounds for his personal preference for 'justice' as a more appropriate term. However, it is clear from his following words that this is not due to the belief that justice provides a settled body of values. Craig wishes only to avoid certain connotations of the rule of law,¹⁵ stating himself to be 'perfectly happy to employ the terminology of the rule of law, provided that it is understood that the substantive concept of the rule of law necessarily entails *some* vision of justice and rights' (p 376, emphasis added). Whose vision of the rule of law (or justice, or rights) is to prevail?

The rule of law is not the only device used on the common law side to cover up judges acting on their own authority to determine the scope of the law. Sir John Laws, in his second contribution to this volume (ch 8), makes use of two devices. He first suggests that the authority of the common law in general and its development of judicial review in particular is based on social consent or public confidence, before abandoning public opinion as 'a many-headed hydra' (p 174). Laws then reaches for the notion of principle with a view to convincing us that it provides a 'process [which] confines the judge's own views in a strict and objective context' (p 189). However, Laws' four building blocks of principle – arguments from logic, from consequences, from precedent, and from ideals – only serve to underscore that principle serves a role subservient to whichever body of doctrine, ideological or otherwise, is selected. For none of these building blocks can provide a definite foundation, nor an architectural plan for the building to be constructed.¹⁶ Principle may be used as a form in which to express the finished product of the judges but it does not provide the substantive authority for their work.

A rather more elaborate device on the common law side is to be found in Dawn Oliver's discussion of a particular set of principles, the 'principles of good administration' (pp 3–4). I need not dwell on the point that what is considered to be good administration is also contentious, and the same point can be made in relation to what may be regarded as a synonymous expression used by Oliver, 'the duties of considerate decision making' (pp 307–308). However, it is important to distinguish between two ways in which these principles or duties can be discussed. It is possible to regard them as providing the authority for judicial decisions, but it is also possible to engage in fruitful discussion under these headings of what factors might be available to be taken into account, and what objectives it might be thought worthwhile to pursue, in formulating principles and in applying them to

15 For Craig's detailed views on the rule of law, see 'Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework' [1997] PL 466.

16 The suggestions that logic provides us with a 'discipline', and that precedent shapes the use of the building blocks into 'a common enterprise' (183), cannot help. The discipline of logic may alert us to what decisions are available, or to what are the consequences of beliefs that we hold, but does not determine which decisions to take or which beliefs to accept. And although precedents once determined may land us all in a particular common enterprise, that is not to say that they have been determined from an enterprise that we have chosen to pursue in common.

cases of judicial review.¹⁷ As valuable as the latter discussion might be, it remains contentious as to how these factors and objectives are to be related and prioritised in particular situations.

Concluding remarks

A common thread has run through the dissatisfaction I have expressed with the 'fairy stories' told on the *ultra vires* side and the 'fig leaves' employed on the common law side. This is their dependence on a non-existent, settled set of values, which could realistically be linked to the principles of judicial review used by the courts – whether as a bridge between parliamentary intention and the principles of judicial review, or in order to provide authorisation for the principles of judicial review.

Something else that can be seen as common to the two models, despite their fierce dispute, is an effort to fix the principles of judicial review within the law. On the *ultra vires* side there is an attempt to subordinate the principles of judicial review to the sovereignty of Parliament, but this is done in a way which does not merely uphold fundamental constitutional doctrine. By tracing the authority for the principles of judicial review used in determining the scope of a particular power to an implied intention regarded as part of the statutory creation of that power, there is retained an appearance of legality to the conditions that the courts impose upon the power.¹⁸ In a very different way, the common law side makes a claim for the legality of these conditions, invoking the rule of law, or some other source of intrinsic legal value. What neither side is prepared to countenance is the prospect that the law enacted by Parliament, or found in the principles of the common law, does not provide an answer to the question: what conditions should attach to the exercise of a power that Parliament has created? If it were admitted that the law did not provide an answer to this question, there could be no pre-existing substantive legal authority for the answer, and hence there would be no subject for a debate focusing on where that authority was to be found.

The defects discovered in the efforts of both sides to provide an answer, and the nature of the disagreement between the two approaches to the basis for judicial review which has bewildered some of the participants in the discussion,¹⁹ both raise the possibility that there is more to what the supporters of the two models are talking about than they know or recognise as lawyers. But lawyers may be reluctant

17 Both roles are apparent in Oliver's work. She treats the principles of good administration as an alternative to *ultra vires* for the basis of judicial review (25–26) but also admits that 'countervailing considerations' may prevail against duties of considerate decision-making (307). Similarly, in 'The Underlying Values of Public and Private Law' in Michael Taggart (ed), *The Province of Administrative Law* (Oxford: Hart Publishing, 1997) 218, she treats 'the key values as keystones', but also perceptively acknowledges that the key values 'have to contend with other considerations in the law and legal policy', at 224. Oliver's concern in these essays, and in her book, *Common Values and the Public-Private Divide* (London: Butterworths, 1999), to identify common underlying abstract values in the law can not provide authority for judicial decisions, otherwise Justinian's identification (*Institutes* 1.3) of the most abstract values of the law – to live honestly, to harm no-one, and to give each his due – would suffice. The problem is that they would suffice to justify a number of conflicting systems of law.

18 Sir William Wade in his Comment sees this as providing an important motive for judges to adopt the *ultra vires* model if only as a matter of expediency. Jeffrey Jowell suggests that just such a motive may have been behind judicial dicta supporting the *ultra vires* model (337).

19 See the Comments of T.R.S. Allan and Michael Taggart. Christopher Forsyth's good humoured self-deprecation in using as an image for the debate the Lilliputian controversy over which end to crack open a boiled egg also testifies to this point (396–397).

to concede that there is not a legal answer, for it implies uncomfortable things for lawyers. One is that if the conditions attached to the exercise of a power granted by Parliament are not in a particular case to be found in the law, it follows that it is not necessarily the job of lawyers to determine them. The second is that if it happens that lawyers are given the job to set these conditions, they should recognise that in doing so they are operating as more than lawyers learned in the law. From this point follows a third uncomfortable implication in raising questions about what skills or qualities we should look for in the lawyers who are handed this task.²⁰

One way of looking at this task emerged in our earlier discussion of reasonableness but can be regarded as characteristic of each of the various abstract depictions referred to as governing this task. It was suggested there that reasonableness as a general objective enjoined a deliberative enterprise rather than a communicative requirement. It was Aristotle who suggested that 'to lay down a law about things that are subjects for deliberation is an impossibility. Therefore men do not deny that it must be for a human being to determine such matters.'²¹ Then Aristotle was not a lawyer.

If these suggestions do have some bearing on the nature of judicial review, the theory of that subject should be concerned not with the authority for the conditions attached to the power that Parliament has granted, but with the issue of whose deliberations beyond the law should determine the scope of the power; in particular, when the scope should be set by the body granted the power, and when by the judges in the courts. This would give prominence to consideration of the role of the judiciary within the theory of judicial review.²²

- 20 Louis Eisenstein, 'Some Iconoclastic Reflections on Tax Administration' (1945) 58 *Harvard Law Review* 477, 523–526, 534–543, argued that judges lacked the abilities to deal with the essentially non-legal issues of working out the details of tax provisions which should be entrusted to an administrative body possessing the appropriate technical skills. That the non-legal aspect of the judicial role should have a bearing on the manner in which judges are appointed has been suggested by John Bell, in ch 10 of *Policy Arguments in Judicial Decisions* (Oxford: Clarendon Press, 1983). More recently, Andrew Fraser, 'Beyond the Charter Debate: Republicanism, Rights and Civic Virtue in the Civil Constitution of Canadian Society' (1993) 1 *Review of Constitutional Studies* 27, has provided a radical suggestion for the election of judges by the legal community as a solution to concerns that an unaccountable judiciary has been given too much power in interpreting the Canadian Charter. John Yoo, 'Choosing Justices: a Political Appointments Process and the Wages of Judicial Supremacy' (2000) 98 *Michigan Law Review* 1436, 1461, 1467, considers, 'By constitutionalizing more areas of life and by pursuing the notion of judicial supremacy, the Court itself has shunted normal political activity from the world of policy into the world of Court appointments.' In order to counter this, he suggests that the appointments process should be used to select new Justices who are prepared to 'deny the Court's own supremacy.'
- 21 Aristotle, *Politics*, III.xi.8. I use the translation from the Loeb edition by H. Rackham (Cambridge, MA: Harvard University Press, 1932) except that I prefer 'determine' to 'judge' in the second sentence. Either translation is appropriate for the Greek *krinonta*, but the former avoids the connotation that we are dealing with a judicial judgment, rather than a human determination. The emphasis on general human determination rather than an exercise of judicial skill is evident from Aristotle's use of *anthropon* here in contrast to the use of *archon* (translated by Rackham as 'official', a term used of the chief magistrates at Athens) in the following passage.
- 22 Treatment of the judicial role has long been a prominent part of discussion of judicial review in the United States. See, eg, Learned Hand, 'Chief Justice Stone's Conception of the Judicial Function' (1946) 46 *Columbia Law Review* 696. That constitutional discourse in the UK may operate to blur our understanding of the judicial role is commented on by Neil Walker, 'Setting English Judges to Rights' (1999) 19 *OJLS* 133, 145. That consideration of the judicial role may gain greater prominence by the enactment of the Human Rights Act 1998 is indicated by suggestions already being made that the practical impact of the Act in a number of areas will depend upon the role performed by the judges. See Helen Fenwick, 'The Right to Protest, the Human Rights Act and the Margin of Appreciation' (1999) 62 *MLR* 491; Gavin Phillipson, 'The Human Rights Act, "Horizontal Effect" and the Common Law: a Bang or a Whimper?' (1999) 62 *MLR* 824; and Stephanie Palmer, 'Human Rights: Implications for Labour Law' (2000) 59 *CLJ* 168.

REVIEWS

Mavis Maclean (ed), Making Law for Families, Oxford: Hart Publishing, 2000, xi + 211 pp, pb £20.00.

Academic family lawyers frequently suggest what they consider to be necessary reforms to the existing body of law. And when they are not recommending reform measures of their own conception, they are often to be found criticising recent legislative developments thought up by others. Much less frequently do they focus their reflective and analytical skills on the actual process of family law making. Maclean's edited volume, *Making Law for Families*, is a collection of essays (originally presented as papers at a workshop convened at Oñati, Spain in 1999) that goes some way towards redressing the balance. It draws on both the practical experiences and perceptive reflections of a fairly impressive array of international academic family lawyers, most of whom – whether as advisors, or as actual draftsmen – have had some involvement in the law-making process. Yet, measured against the objective set by the editor that this volume should help 'to develop the theory of the legislative process in a grounded way' (p 1), the collection warrants only a lukewarm reception.

While each of the nine substantive contributions makes for interesting reading, it is doubtful whether all them contribute to the editor's stated objective. Take, for example, Benoit Bastard's 'Administrative Divorce in France: A Controversy Over a Reform, that Never Reached the Statute Book'. This essay provides little more than a descriptive account of a reform proposal that fell by the wayside in France largely because of institutional opposition. In documenting the rise and fall of the proposal – in essence, that the French divorce process should be conducted in town halls not courts, and be managed by mayors not judges – Bastard's essay relies far too heavily on the simple *bloc* quotation of large chunks of articles that appeared in a select few French newspapers. Indeed, roughly six pages of this twenty-page contribution are taken up in this way, and the cost to the chapter is levied in terms of academic merit. For while the essay clearly demonstrates that family law reform in France has become as politicised as in this and other countries, the point could have been made much more succinctly. In a frank admission, the author himself confesses that in reading his essay 'we are doing no more than considering an example of a projected legal reform which ... has not been adopted, and is not for the time being on the statute book' (p 72).

Chapters 7 and 8 are similarly descriptive essays which provide accounts of family policy and family law making in Poland and Bulgaria since 1989 and 1997 respectively. Here, again, while the contributions make for interesting reading, they do so more in the nature of modern legal history than insightful comparative family law. To take one example from the chapter by Malgorzata Fuszara and Beata Laciak, we are informed that the divorce rate in Poland in 1989 was a shade over 1 per 1,000 of the population (compared with nearly 13 per 1,000 in this country at that time). We are then told that from 1994 there had been 'a slow but systematic [*sic*] increase in the number of divorces in Poland' (p 120) such that, by 1997, the rate had crept up to a dizzying 1.1 per 1,000. The following five pages then describe (somewhat repetitiously) the campaign to reform the law so as to curb this 'alarming' increase in the divorce rate, only to inform us at the very end of the section that '[t]he 1997 Parliamentary elections put a stop to work on the draft Bill' (p 124). Again, the only

real lesson that we draw from this essay is that family law has become intensely politicised in Poland.

A no less enjoyable read, but one which, again, has precious few lessons for comparative family lawyers is Encarna Roca's 'Regulation of Same-Sex Partnerships from a Spanish Perspective'. Although same-sex relationships are an ongoing family law concern in many countries, the truth is that Roca's contribution deals with the subject from such an esoteric and narrow perspective that what it teaches us is of less value than we might initially have been forgiven for thinking. The essay is centred on the constitutional restrictions placed upon the legislative competencies of the Spanish *Autonomias* of Catalonia and Aragon (with respect to affording a kind of quasi-marital status to stable, long-term same-sex relationships). It demonstrates little more than that these autonomous Communities have legislated on this matter primarily to secure the political goal of 'seeking to establish their place among the community of "progressive" countries' (p 95). Thus, while Roca's essay endorses the point that surfaces time after time in this volume – that legislating on family law matters (whether in this country or abroad) has acquired an unquestionably political dimension in recent years – it is not clear what else those interested in the generalities of the law-making process (or, indeed, the appropriate status for same-sex relationships) are to garner from such a subject-specific chapter.

Happily, the collection has its strong points, too. An altogether more penetrating and stimulating discussion of the politicisation of family law – the 'new political economy of family law', as the author puts it – is to be found in John Dewar's essay, 'Making Family Law New? Property and Superannuation Reform in Australia'. He begins by locating family law reform generally within a broader policy objective of matching modern family laws to modern patterns of living (while at the same time increasing the transparency and efficiency of the family law system). He then demonstrates that the perceptible shift away from a 'technocratic' family law system – one characterised by 'a belief in the power of specialist institutions, official expertise and state-sanctioned discretion to deal with the problems of family adjustment' (p 63) – amounts to a significant reconfiguration of family law which itself has important implications. In particular, Dewar identifies 'new normative types, new sources of norms, new audiences for legislation and a self-conscious attempt to address the fragmentation of the family law system' (p 69).

In similar vein, John Eekelaar's essay, 'Uncovering Social Obligations: Family Law and the Responsible Citizen', points towards a shift in the normative basis of family law before suggesting that we ought to 're-characterise the domain of family law as the law of personal rights and obligations' wherein the relevant 'rights and obligations ... include social as well as legal ones' (p 27). Although there is some engagement with familiar positivist contentions about the nature of, and distinctions between, legal and social norms, Eekelaar ultimately concludes that there is a certain redundancy to that debate in the present context: 'Legal theorists closer to the work of sociologists and anthropologists ... have asserted that law can arise from social practices as well as formal institutions and it is difficult to resist this from a system of family and kinship systems' (p 14).

For Eekelaar, the principle that fathers should pay for their children has always been in the domain of family law despite the fact that the legislative instrument giving this principle formal legal pedigree did not emerge until 1991 in the form of the Child Support Act. According to Eekelaar's account, the child-support legislation did not so much implement a new family law norm as 'attempt to strengthen what was perceived to be the weakening social norm that fathers should pay for their children' (p 19).

Taken together, Dewar's 'reconfiguration of family law' and Eekelaar's reconceptualisation of the same contain much for judges, legislators and academic commentators. So too, in its own way, does Mary Slaughter's essay which applies game theory to common modes of marital and family life with a view to offering new insights to legislators. In short, she endeavours to show that family law can be used as a conduit through which new norms may be channelled in order to create a change in the sense of respective entitlements of husbands and wives. For Slaughter, everything turns on the distribution of the marital surplus – that is, the greater degree of welfare produced for both parties *because of* their marriage. This distribution, she explains, 'will depend on the relative strength of each party to drive a hard bargain' (p 39). Thus, noting (at p 45) that there are 'structural inequalities in the public world which tend to cause inequalities in the private world', she suggests that such asymmetries can best be addressed if:

[through law] more equitable arrangements could be struck *ex ante* – before the marriage begins and the children are born – either through private ordering or through the law's default rules ... [for] it would solidify the wife's position at the beginning and stop the downward spiral that occurs because her bargaining power decreases if the marriage becomes more important to her than her husband (p 45).

It is a pity that she leaves the matter hanging in this way. In a climate in which the trend has been steadily towards a mere residual role for substantive family law, it is at the very least curious that she fails to differentiate between, or reflect upon the *true* potential for the 'private ordering' and 'default rules' solutions she offers.

The collection closes with two essays that focus on the human rights dimension. In the wake of the Human Rights Act 1998 it is hardly surprising that some such contributions should have been included. Both Clare Archbold's 'Family Law-Making and Human Rights in the United Kingdom' and Anne Griffiths' and Randy Kandel's 'Legislating for the Child's Voice' are welcome contributions to the rapidly burgeoning literature in this field. The former is particularly worthy of note given the author's position at the Northern Ireland Office of Law Reform.

On the whole, *Making Law for Families* is a worthwhile book, despite it being a curate's egg. A final quibble is that, quite apart from the dubious quality of some of the contributions, it has no shortage of other little irritations such as an uncommonly high number of typographical slips and errors of grammar and syntax. While some of the latter ought perhaps to be forgiven on the basis that English is not the mother tongue of the authors concerned, it is unforgivable that so many inconsistencies in referencing style – not only between, but also within, contributions – should have been allowed to slip through the net. For the most part, the Harvard system seems to have been adopted. This, of course, is fine in itself; but on several occasions the reader keen to follow through a reference will turn the relevant pages only to discover the annoying absence of a journal volume number or a missing page reference. There are quite a number of such 'blips' throughout the volume, but by far the most amusing one appears in Roca's essay: 'full relevance to be supplied at proof'.

John Murphy*

* University of Manchester.

Paul W. Kahn, *The Cultural Study of Law: Reconstructing Legal Scholarship*, Chicago: University of Chicago Press, 1999, ix + 169pp, hb £20.00.

'The issue is not whether law makes us better off, but rather what it is that law makes us' (p 6). This is Paul Kahn's central refrain in *The Cultural Study of Law*. Too little theoretical attention is given to the facticity of law: there is plenty of legal scholarship, of course, but it is mostly tied up with legal practice. The ruling assumption, according to Kahn, is that everything studied in a faculty of law has to be somehow useful to lawyers. By trying to provide a service for lawyers, the legal scholar becomes a sort of cheerleader, committed to making law work (or work better). Kahn wants to find a space for agnosticism in legal scholarship, and is trying to demonstrate what the agnostic legal scholar would wish to study.

Kahn's project is not anti-theoretical: he is not belittling the ambitions of the normative legal scholar or speculating as to the absent consequences of normative theory. Nor is he seeking to write that sort of theory out of the academy. He is merely trying to establish room for what he sees as 'a necessary supplement to our current practice of legal studies' (p 40). In fact, his complaint is as much about style as about content. He does not maintain that all scholars of law fail (although some undoubtedly do fail) to investigate the phenomenology of law's rule. His point, rather, is that all such scholars conclude their studies in a programmatic fashion, thereby implicitly denying the possibility that the study of law might be a good in itself.

This book is not itself a cultural study of law, but an argument for the value of such study and an explanation of the conditions of that study. From the outset of the study we encounter the notion of the rule of law. We subscribe to the order of law. But this order is not 'out there' as some sort of objective truth. Rather, it is constituted through our beliefs. Our commitment to the rule of law is not a product of some metaphysical schema, but of our own imagination. Typical normative studies of law present the rule of law in terms of some distinct social or moral vision rather than as a *mélange* of such visions. Now, it might be objected here that Kahn's position is a distinctly conservative one. He seems to be valorising the rule of law, insisting that we do not tamper with it. But in fact his point is rather more subtle. In part, he is contending that the law is not a product of a systematic rationality and cannot be developed by a systematic rationality. There exists a number of competing goals and values, and it is inappropriate, indeed impossible, to tie law to just one of them. He is also arguing that there may be something quite special and useful about law's rule. It may be exactly because law works through bricolage that it works for us. Clearly we must identify and work on the problems with law's rule, but there is no intellectual gain to be had from attempting to tag the law to a particular moral order.

Kahn seeks a fairly straightforward appraisal of what it means to live under the rule of law. Such an appraisal would involve contrasting the order of law with the orders of love and nature and revolution. Law may be less fulfilling, less erotic than those alternative orders, but then what does law's difference from those alternative orders tell us about ourselves? Kahn is keen to emphasise the interrogative potential of the cultural study of law. One example of how he does this (drawn from pp 110–12) might be highlighted. Kahn complains of how international lawyers see the proliferation of treaty law and the formation of international 'supernorms' as a progressive development. However, the problem with this 'myth of progress' is that it obscures the fact that the threat of violent destruction has become 'a global background condition of everyone's existence'. For Kahn, the central question in international law is about 'the relationship between the frenzy of law-making activity

over the past one hundred years and the recourse to unconstrained violence that marks so many of this century's conflicts'. He speculates, for instance, that it is perhaps exactly because genocide is the paradigm of a *jus cogens* norm that it has been incorporated into the agenda of many regimes. The development of international law sets boundaries of behaviour that invite transgression as well as compliance. It is only if we investigate either response that we can hope to understand what it is to live under law. And it is only if we rid ourselves of the myth of progress that we can hope to study fully the phenomenology of law.

On the other hand, it may also be essential to the cultural study of law that we realise that law provokes normative speculation. Is not the desire for justice one of the very reasons we do subscribe to the order of law? The problem for Kahn may be that the myth of progress is integral to legal order. Indeed, one critical response to Kahn's thesis may be that just as the normative legal scholar assumes that law is 'a failed form of something other than itself', Kahn assumes that it is not. He does, none the less, seem to have a point in arguing that there is a realm of meanings and a set of questions that is being neglected by normative legal scholarship. He may well be making and hiding assumptions of his own as he goes about the task of exposing the assumptions of others, but then he is not suggesting that we should all become Kahnians. He does, after all, pose the cultural study of law only as a necessary supplement to other types of legal scholarship.

Emran Mian*

Rex J. Ahdar (ed), **Law and Religion**, Ashgate: Dartmouth, 2000, xiii + 244pp, hb £55.00.

In many respects, this is a very important book. What is impressive in this particular work is the uniform consistency of quality in all of its content, a characteristic that is notably lacking in many collections of solicited articles. The book begins with an introduction by Rex Ahdar on the 'inevitability of law and religion', a somewhat cryptic phrase. The content is nevertheless clear, and useful as a beginning. Ahdar demonstrates that, if law and religion stand and fall together, the analysis of an interaction between them is hardly an unimportant task. If we agree with Ahdar, we discover that secularisation theory failed to dismiss religion as a withering force and, instead, a religio-legal interdependence was fostered.

The second chapter in the book under review is an intriguing essay by Calum Carmichael on the theme of in what way the ten commandments are 'religious'. We infer from Carmichael that every legal system needs some supernatural imprimatur to lend authority to its contents; using his terms, a divine imprimatur is a mask that can enhance the law's efficiency. It remains an open question, though, whether the mythological dimension to law itself is practically useful. The third essay of the volume could not contrast more markedly with the second. This is a programmatic and pragmatic contribution from Malcolm Evans on the work of the Human Rights Committee in terms of the United Nations and freedom of religion. He gives a rather sobering picture of a world rife with religious persecution, and considers three questions concerning the Committee's approach – what counts as a religion for the

* Corpus Christi College, Cambridge.

purposes of protection; what are religious manifestations or actions; and, when are state restrictions upon religious conduct justified. These three elements interplay, and the Committee handles them with difficulty and, we deduce, with undue caution. This conservative stance is, in general, unhelpful and ineffective. The Committee needs revitalisation.

The book's fourth essay is by Michael McConnell on tensions in American First Amendment doctrine. McConnell is one of the United States' foremost legal authorities on the two religious clauses of the First Amendment. He illuminates here First Amendment church-state jurisprudence and examines exemptions from the general law of the land for religionists along with government fiscal assistance to religiously-based bodies. He is frustrated with the lack of understanding of the true purpose of the religious clauses, which has led to inconsistency in their interpretation. In the fifth essay, Marie Failinger continues the church-state critique with special emphasis on the Supreme Court. It illuminates Stanley Fish's characterisation of the just bounds between church and state as a 'mission impossible'. To understand the complicated case law we must examine the problems of power, solicitude for individual freedom and the conflict of opposing ideologies. The problem of ideology is highlighted as the key explanatory idea, and this is stimulating. It is not simply a disguised restatement of the recurrent American theme of the diffusion of power to counteract corruption and tyranny. The American trilogy of essays that forms the central pillar of *Law And Religion* is rounded off by James Richardson's chosen approach to the religious clauses. He focuses on cults and injustices done to them, recommending a healthy scepticism of expert evidence and an emphasis on solid proof only in relation to them. These matters are examined with consummate expertise and authority, and Richardson's views should be treated as highly pertinent and influential.

With the sixth essay in the book under review, we find new terrain. Attention switches to the United Kingdom, and its newly-operational Human Rights Act 1998. Here Julian Rivers covers some surprisingly original ground, voicing understated opinions based on a very sound understanding of religion in general. Remaining in Europe for the seventh essay, Sophie van Bijsterveld informs us intelligently about existing European rights-systems. This is a refreshing, deceptively optimistic analysis, revealing the promise in current European developments whatever their provenance and official value. In essay nine, Reid Mortensen gives us nothing less than a *tour de force*. His essay, which draws upon Australasian scholarship, reveals blasphemy's uncomfortable place in secular liberal democracies. Whilst a certain artistic sovereignty and concomitant literary licence are important, Mortensen refuses to ignore the viability of genuine expressions of offence regarding the rather cavalier artistic treatment of various convention-determined sacred symbols. A more avant-garde approach would posit that art invariably reinvigorates such symbols by transmuting them into the artistic dimension, which would be in accordance with the desires of a truly creative god.

The final essay in *Law And Religion* is narrow in its subject and scope. Here Davina Cooper gives a sociological account of Barnett's *eruv*. This local analysis contrasts very abruptly with the more general themes presented by the other contributors. Nevertheless, it covers the much-neglected area of the rightful place of Jewish religio-cultural iconography in contemporary Britain. It reveals immanent anti-semitism in Britain in a deceptively discreet way, and the detailed analysis shows a local dispute with potentially very broad negative ramifications in terms of social relations. What seems *prima facie* a somewhat misplaced contribution becomes the main flag-bearer for religious equality.

Law And Religion is a very accomplished aggregate of inspiring ideas on a theme deserving of serious attention. As with all such collective works, there are some contributions that are more impressive than others, but it contains contributions of a near-even good quality. I would draw attention, though, to a few matters of surprising absence. It is strange for a book of this kind to ignore completely the concept of natural law, especially, in this work, theological natural law. Nor does the text entertain theoretical observations directly on the relationship between law and morality. Buddhism and Hinduism are neglected, as are the legitimate concerns of atheists and agnostics that religious labels, *per se*, have caused much unnecessary hostility. Christian Science and the faiths of Jehovah Witnesses and Mormons also do not feature. Nevertheless, this book demonstrates that it is important for law to consider the soul and its welfare in an era when faith and true *oeuvres d'esprit* are comparatively rare and frequently conceptually derided.

Paul Kearns*

* School of Law, University of Manchester.

Information for Contributors

Contributions to and correspondence concerning the various sections of the Review should be sent to the relevant Editor:

Articles: Hugh Collins, Law Department, London School of Economics, Houghton Street, London WC2A 2AE

Legislation and Reports: R.C. Simpson, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Case Notes: Colin Scott, Law Department, London School of Economics, Houghton Street, London WC2A 2AE.

Books for Review, Review Articles and Reviews: Neil Duxbury, Faculty of Law, University of Manchester, Oxford Road, Manchester M13 9PL.

General Correspondence: Tim Murphy, Law Department, London School of Economics, Houghton Street, London WC2A 2AE

Section Editors will be happy to discuss drafts and proposed contributions with contributors. All contributions should be typed, double spaced and include an estimate of the number of words, including footnotes. Articles should be preceded by a short abstract of approximately 100–150 words. Reports, Legislation and Case Notes should be preceded by a list of all reports, legislation and cases cited in the paper, displayed as keywords. Contributions should use footnotes which make cross-references within the text as sparingly as possible. Contributions must conform to the Review's house style; a style sheet is available on the MLR website or will be provided by the Production Editor on request.

Articles do not usually exceed 12,000 words (excluding footnotes) Longer or shorter articles can be considered.

Subscription Information 2001

The Modern Law Review is published six times a year. Subscription rates are as follows:

Annual subscription for unbound parts only

UK and Europe	£108.00
Rest of the world	£114.00
The Americas	US\$186.00*
UK and European law teachers and law students	£25.00
Commonwealth law teachers and law students	£32.00/US\$50.00*

Annual subscriptions for Bound Volume Service unbound parts plus a bound volume

UK and Europe	£159.00
Rest of the world	£169.00
The Americas	US\$266.00*
UK and European law teachers and students	£65.00
Commonwealth law teachers and law students	£68.00/US\$109.00*

Back Issues

Single issues of the unbound parts of the last three volumes are available from the publishers at the current price for unbound parts:

	£22.00 (£23.00 for ROW, US\$37.00 The Americas)*
Single bound volumes	£71.00 (£74.00 for ROW, US\$122.00 The Americas)*
Complete sets of bound volumes	£2071.00 (US\$3417.00 The Americas)*

*Canadian customers/residents please add 7% GST.

Orders, sample copy requests, claims and other subscription enquiries should be addressed to: Blackwell Publishers Journals, PO Box 805, 108 Cowley Road, Oxford OX4 1FH, UK. All subscriptions are supplied on a calendar year basis (January to December).

US Mailing

Periodicals postage paid at Rahway, New Jersey.

Postmaster: send address corrections to *The Modern Law Review*, c/o Mercury Airfreight International Ltd Inc., 365 Blair Road, Avenel, NJ 07001, USA (US mailing agent).

Microform

The journal is available on microfilm (16 mm or 35 mm) or 105 mm microfiche from the Serials Acquisitions Department, Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106, USA.